

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 27, 2001, 15:40:56 ; Search time 8748.63 Seconds
(without alignments)
3712.172 Million cell updates/sec

Title: US-09-715-036-4
Perfect score: 1853
Sequence: 1 atggctaagctcttgagctt.....aaaaaaaaaaaaaaaaaaaaa 1853

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 17159718 seqs, 876320856 residues 34319436

Total number of hits satisfying chosen parameters: 34319436

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :	Pending Patents_NA_Main:*
1:	/cgn2_6/ptodata/2/pna/US06111_COMB.seq:*
2:	/cgn2_6/ptodata/2/pna/US06112_COMB.seq:*
3:	/cgn2_6/ptodata/2/pna/US07_COMB.seq:*
4:	/cgn2_6/ptodata/2/pna/US080_COMB.seq:*
5:	/cgn2_6/ptodata/2/pna/US081_COMB.seq:*
6:	/cgn2_6/ptodata/2/pna/US082_COMB.seq:*
7:	/cgn2_6/ptodata/2/pna/US083_COMB.seq:*
8:	/cgn2_6/ptodata/2/pna/US084_COMB.seq:*
9:	/cgn2_6/ptodata/2/pna/US085_COMB.seq:*
10:	/cgn2_6/ptodata/2/pna/US086_COMB.seq:*
11:	/cgn2_6/ptodata/2/pna/US087_COMB.seq:*
12:	/cgn2_6/ptodata/2/pna/US088_COMB.seq:*
13:	/cgn2_6/ptodata/2/pna/US089_COMB.seq:*
14:	/cgn2_6/ptodata/2/pna/US090_COMB.seq:*
15:	/cgn2_6/ptodata/2/pna/US091_COMB.seq:*
16:	/cgn2_6/ptodata/2/pna/US092_COMB.seq:*
17:	/cgn2_6/ptodata/2/pna/US093_COMB.seq:*
18:	/cgn2_6/ptodata/2/pna/US094_COMB.seq:*
19:	/cgn2_6/ptodata/2/pna/US095A_COMB.seq:*
20:	/cgn2_6/ptodata/2/pna/US095B_COMB.seq:*
21:	/cgn2_6/ptodata/2/pna/US095C_COMB.seq:*
22:	/cgn2_6/ptodata/2/pna/US095D_COMB.seq:*
23:	/cgn2_6/ptodata/2/pna/US096A_COMB.seq:*
24:	/cgn2_6/ptodata/2/pna/US096B_COMB.seq:*
25:	/cgn2_6/ptodata/2/pna/US096C_COMB.seq:*
26:	/cgn2_6/ptodata/2/pna/US096D_COMB.seq:*
27:	/cgn2_6/ptodata/2/pna/US096E_COMB.seq:*
28:	/cgn2_6/ptodata/2/pna/US097A_COMB.seq:*
29:	/cgn2_6/ptodata/2/pna/US097B_COMB.seq:*
30:	/cgn2_6/ptodata/2/pna/US097C_COMB.seq:*
31:	/cgn2_6/ptodata/2/pna/US098_COMB.seq:*
32:	/cgn2_6/ptodata/2/pna/US099_COMB.seq:*
33:	/cgn2_6/ptodata/2/pna/US6000_COMB.seq:*
34:	/cgn2_6/ptodata/2/pna/US6001_COMB.seq:*
35:	/cgn2_6/ptodata/2/pna/US6002_COMB.seq:*
36:	/cgn2_6/ptodata/2/pna/US6003_COMB.seq:*
37:	/cgn2_6/ptodata/2/pna/US6004_COMB.seq:*
38:	/cgn2_6/ptodata/2/pna/US6005_COMB.seq:*
39:	/cgn2_6/ptodata/2/pna/US6006_COMB.seq:*
40:	/cgn2_6/ptodata/2/pna/US6007_COMB.seq:*
41:	/cgn2_6/ptodata/2/pna/US6008_COMB.seq:*
42:	/cgn2_6/ptodata/2/pna/US6009_COMB.seq:*
43:	/cgn2_6/ptodata/2/pna/US6010_COMB.seq:*

44:	/cgn2_6/ptodata/2/pna/US6011_COMB.seq:*
45:	/cgn2_6/ptodata/2/pna/US6012_COMB.seq:*
46:	/cgn2_6/ptodata/2/pna/US6013_COMB.seq:*
47:	/cgn2_6/ptodata/2/pna/US6014_COMB.seq:*
48:	/cgn2_6/ptodata/2/pna/US6015_COMB.seq:*
49:	/cgn2_6/ptodata/2/pna/US6016_COMB.seq:*
50:	/cgn2_6/ptodata/2/pna/US6017_COMB.seq:*
51:	/cgn2_6/ptodata/2/pna/US6018_COMB.seq:*
52:	/cgn2_6/ptodata/2/pna/US6019_COMB.seq:*
53:	/cgn2_6/ptodata/2/pna/US6020_COMB.seq:*
54:	/cgn2_6/ptodata/2/pna/US6021_COMB.seq:*
55:	/cgn2_6/ptodata/2/pna/US6022_COMB.seq:*
56:	/cgn2_6/ptodata/2/pna/US6023_COMB.seq:*
57:	/cgn2_6/ptodata/2/pna/US6024_COMB.seq:*
58:	/cgn2_6/ptodata/2/pna/US6025_COMB.seq:*
59:	/cgn2_6/ptodata/2/pna/US6026_COMB.seq:*
60:	/cgn2_6/ptodata/2/pna/US6027_COMB.seq:*
61:	/cgn2_6/ptodata/2/pna/US6028_COMB.seq:*
62:	/cgn2_6/ptodata/2/pna/US6029_COMB.seq:*
63:	/cgn2_6/ptodata/2/pna/US6030_COMB.seq:*
64:	/cgn2_6/ptodata/2/pna/US6031_COMB.seq:*
65:	/cgn2_6/ptodata/2/pna/US6032_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1853	100.0	1853	28	US-09-715-036-4	Sequence 4, Appli
2	1391.2	75.1	1524	15	US-09-141-220-5	Sequence 5, Appli
3	1391.2	75.1	1524	15	US-09-141-220B-5	Sequence 5, Appli
4	1391.2	75.1	1524	15	US-09-141-220C-5	Sequence 5, Appli
5	1391.2	75.1	1524	18	US-09-494-096-6	Sequence 6, Appli
6	1377.4	74.3	1524	18	US-09-455-294-7	Sequence 7, Appli
7	527.2	28.5	1446	13	US-08-913-588-13	Sequence 13, Appl
8	527.2	28.5	1446	15	US-09-108-010-13	Sequence 13, Appl
9	527.2	28.5	1446	29	US-09-758-652-13	Sequence 13, Appl
10	516	27.8	1743	30	US-09-762-381-1	Sequence 1, Appli
11	515.6	27.8	1488	13	US-08-913-588-11	Sequence 11, Appl
12	515.6	27.8	1488	15	US-09-108-010-11	Sequence 11, Appl
13	515.6	27.8	1488	29	US-09-758-652-11	Sequence 11, Appl
14	510.8	27.6	1693	64	US-60-312-544-1040	Sequence 1040, Ap
15	508.2	27.4	1458	13	US-08-913-588-12	Sequence 12, Appl
16	508.2	27.4	1458	15	US-09-108-010-12	Sequence 12, Appl
17	508.2	27.4	1458	29	US-09-758-652-12	Sequence 12, Appl
18	355.8	19.2	1323	25	US-09-654-617-258258	Sequence 258258,
19	355.8	19.2	1323	27	US-09-684-016-258258	Sequence 1113, Ap
20	352	19.0	1230	64	US-60-312-544-1113	Sequence 199, App
21	333.2	18.0	1801	25	US-09-654-617-199	Sequence 199, App
22	333.2	18.0	1801	27	US-09-684-016-199	Sequence 198, App
23	303	16.4	1830	25	US-09-654-617-198	Sequence 188, App
24	303	16.4	1830	27	US-09-684-016-198	Sequence 18985,
25	220.2	11.9	1855	25	US-09-654-617-189985	Sequence 189985,
26	220.2	11.9	1855	27	US-09-684-016-189985	Sequence 15, Appl
27	201.8	10.9	1551	13	US-08-913-588-15	Sequence 15, Appl
28	201.8	10.9	1551	15	US-09-108-010-15	Sequence 15, Appl
29	201.8	10.9	1551	29	US-09-758-652-15	Sequence 15, Appl
30	201.6	10.9	3527	29	US-09-759-967-1	Sequence 1, Appli
31	197.6	10.7	416	17	US-09-304-517A-253748	Sequence 253748,
32	197.6	10.7	416	17	US-09-371-146A-253748	Sequence 253748,
33	197.6	10.7	416	18	US-09-440-687-15940	Sequence 15940, A
34	197.6	10.7	416	47	US-60-144-084-4016	Sequence 4016, Ap
35	196	10.6	395	17	US-09-304-517A-255598	Sequence 255598,
36	196	10.6	395	17	US-09-371-146A-255586	Sequence 255586,
37	196	10.6	395	22	US-09-565-240-2017	Sequence 2017, Ap
38	196	10.6	395	47	US-60-141-135A-1389	Sequence 1389, Ap
39	194.4	10.5	362	17	US-09-304-517A-233545	Sequence 233545,
40	194.4	10.5	362	17	US-09-371-146A-233545	Sequence 233545,
41	194.4	10.5	362	17	US-09-391-630-1117	Sequence 1117, Ap

QY	1	atggctaagctctcttgagctctctctttctgtctcttctgtctcttggtgagctagcagc	60
Db	1	atggctaagctctcttgagctctctctttctgtctcttctgtctcttggtgagctagcagc	60
QY	61	atctcttcaggcagcagcgagaggaatgcgtccaggttcacgagcctcaatgcgcag	120
Db	61	atctcttcaggcagcagcgagaggaatgcgtccaggttcacgagcctcaatgcgcag	120
QY	121	agacctgacaaaccgattgaatcggaggcggttacattgagacttgaaacccccaaacac	180
Db	121	agacctgacaaaccgattgaatcggaggcggttacattgagacttgaaacccccaaacac	180
QY	181	caggagttcgaattgcgcggcgctgcgcctctctcgtcttagtcctccgcgcgaagccctt	240
Db	181	caggagttcgaattgcgcggcgctgcgcctctctcgtcttagtcctccgcgcgaagccctt	240
QY	241	cgtaggcctttctactccaatgcgtccccaggagatcttctatccagcaaggaggatcac	300
Db	241	cgtaggcctttctactccaatgcgtccccaggagatcttctatccagcaaggaggatcac	300
QY	301	tttggtgtgatcttccctggtgtctctagcacatagaagagctgcacacaagcagc	360
Db	301	tttggtgtgatcttccctggtgtctctagcacatagaagagctgcacacaagcagc	360
QY	361	cgatatcagtcaccaagaccacccaagacgtttgcagaagaagaccaaagccacacgcaa	420
Db	361	cgatatcagtcaccaagaccacccaagacgtttgcagaagaagaccaaagccacacgcaa	420
QY	421	caaatagtcacagaaggtgcacogttccaatgaggtgatctcattgcagttccacc	480
Db	421	caaatagtcacagaaggtgcacogttccaatgaggtgatctcattgcagttccacc	480
QY	481	gggtgtgcttctcgctgtacacagccacacactgatgttgtgtgtcttcttact	540
Db	481	gggtgtgcttctcgctgtacacacacacacacactgatgttgtgtgtcttcttact	540


```

; APPLICANT: Burks Jr., A. Wesley
; APPLICANT: Sampson, Hugh A.
; APPLICANT: Sosin, Howard B.
; TITLE OF INVENTION: Methods and Reagents for Decreasing Clinical Reactions
; TITLE OF INVENTION: To Allergy
; FILE REFERENCE: HS 102 CIP (2)
; CURRENT APPLICATION NUMBER: US/09/494,096
; CURRENT FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: 09/141,220
; PRIOR FILING DATE: 1998-08-27
; PRIOR APPLICATION NUMBER: 09/240,557
; PRIOR FILING DATE: 1999-01-29
; PRIOR APPLICATION NUMBER: 09/241,101
; PRIOR FILING DATE: 1999-01-29
; PRIOR APPLICATION NUMBER: 09/248,673
; PRIOR FILING DATE: 1999-02-11
; PRIOR APPLICATION NUMBER: 60/073,283
; PRIOR FILING DATE: 1998-01-29
; PRIOR APPLICATION NUMBER: 60/074,590
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: 60/074,624
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: 09/248,674
; PRIOR FILING DATE: 1999-02-11
; PRIOR APPLICATION NUMBER: 60/122,566
; PRIOR FILING DATE: 1999-03-02
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 1524
; TYPE: DNA
; ORGANISM: Arachis hypogaea
US-09-494-096-6

```

```

Query Match      75.1%; Score 1391.2; DB 18; Length 1524;
Best Local Similarity 95.4%; Pred. No. 1.8e-230;
Matches 1454; Conservative 0; Mismatches 68; Indels 2; Gaps 2;

QY 71 ggcagcagccgagagagatcgctccagctccagcgcctcaatcgagacacctgaca 130
DB 2 ggcagcaacccgagagagacgctgctccagctccagcgcctcaatcgagacacctgaca 61
QY 131 accgcatgaatcgagggcggttaccatgagacttggaaccccaacacccagggatcg 190
DB 62 atcgcatgaatcgagggcggttaccatgagacttggaaccccaacacccagggatcg 121
QY 191 atgcccggcgctgcgcctctctctctctctctctctctctctctctctctctctct 250
DB 122 atgcccggcgctgcgcctctctctctctctctctctctctctctctctctctctct 181
QY 251 tctactccaatgctcccccagagatctctatccagcaagaaggagatactttggttga 310
DB 182 tctactccaatgctcccccagagatctctatccagcaagaaggagatactttggttga 241
QY 311 tctctcctggttctctagacataatgaagagcctgacacaaagagcccgatcagtt 370
DB 242 tctctcctggttctctagacataatgaagagcctgacacaaagagcccgatcagtt 301
QY 371 cccaagacacccaagacgcttgcgaagagagagacacaaagcccaacagacagatagtc 430
DB 302 cccaagacacccaagacgcttgcgaagagagagacacaaagcccaacagacagatagtc 361
QY 431 accagagatgaccccttcaatgaggtgactctcaattgacttccaccggtgtgctt 490
DB 362 accagaggtgaccccttcaatgaggtgactctcaattgacttccaccggtgtgctt 421
QY 491 tctggctgtacacacacacacacacacacacacacacacacacacacacacacacacac 550
DB 422 tctggctgtacacacacacacacacacacacacacacacacacacacacacacacacac 481
QY 551 acaacgacacacacacacacacacacacacacacacacacacacacacacacacacacac 610

```

```

DB 482 acaacgacacacacacacacacacacacacacacacacacacacacacacacacacac 541
QY 611 aagagttcttaagttaccagcagacacacacacacacacacacacacacacacacacacac 670
DB 542 aagagttcttaagttaccagcagacacacacacacacacacacacacacacacacacacac 601
QY 671 gccatatacagcccgcatagtcgctccagacgagacgagacgagacgagacgagacgagac 730
DB 602 gccatatacagcccgcatagtcgctccagacgagacgagacgagacgagacgagacgagac 661
QY 731 agcacagccagacgagacgagacgagacgagacgagacgagacgagacgagacgagacgag 790
DB 662 agcacagccagacgagacgagacgagacgagacgagacgagacgagacgagacgagacgag 721
QY 791 gcggttcacgcccggagttcctggacacacacacacacacacacacacacacacacacacac 850
DB 722 gcggttcacgcccggagttcctggacacacacacacacacacacacacacacacacacacac 781
QY 851 aaaatctgtggcgagacgagacgagacgagacgagacgagacgagacgagacgagacgagac 910
DB 782 aaaacctaagagcgagacgagacgagacgagacgagacgagacgagacgagacgagacgag 841
QY 911 gcctcagaatcttgagcccagatggaacgagacgagacgagacgagacgagacgagacgag 970
DB 842 gcctcagaatcttgagcccagatggaacgagacgagacgagacgagacgagacgagacgag 901
QY 971 aagatcaatataatgaatgaacgagacgagacgagacgagacgagacgagacgagacgag 1030
DB 902 aagatgaatataatgaatgaacgagacgagacgagacgagacgagacgagacgagacgag 961
QY 1031 ggaatggtattgaagagacgagacgagacgagacgagacgagacgagacgagacgagacgag 1090
DB 962 ggaatggtattgaagagacgagacgagacgagacgagacgagacgagacgagacgagacgag 1021
QY 1091 gatccctcaatctacgacatccctcagcgcgtggttcaactcaaaactgcc-acgactccaac 1149
DB 1022 gatccctcagatctacacacacacacacacacacacacacacacacacacacacacacac 1080
QY 1150 ctttaactccttagtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtg 1209
DB 1081 ctttaactccttagtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtg 1140
QY 1210 ttgttttccctcactcaacacacacacacacacacacacacacacacacacacacacacac 1269
DB 1141 ttgttttccctcactcaacacacacacacacacacacacacacacacacacacacacacac 1200
QY 1270 gctcagctgcaagtgtggtggtggtggtggtggtggtggtggtggtggtggtggtggtg 1329
DB 1201 gctcagctgcaagtgtggtggtggtggtggtggtggtggtggtggtggtggtggtggtg 1260
QY 1330 ggtcacgttctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1389
DB 1261 ggtcacgttctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1320
QY 1390 ttgcaatacgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1449
DB 1321 ttgcaatacgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1380
QY 1450 aactcctcctatagataaactcctcctcctcctcctcctcctcctcctcctcctcctcct 1509
DB 1391 aactcctcctatagataaactcctcctcctcctcctcctcctcctcctcctcctcctcct 1440
QY 1510 gagcagggcagggcagggcagggcagggcagggcagggcagggcagggcagggcagggcag 1569
DB 1441 gagcagggcagggcagggcagggcagggcagggcagggcagggcagggcagggcagggcag 1500
QY 1570 cagtctccgagggcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1593
DB 1501 cagtctccgagggcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1524

```


Qy	1333	caogttcttgggtgcacagaacttcg	ccgtgggtgggaagtc	ccagagcagaacttc	1392
Db	1186	CAGGTGTTAAATGTGCGCACAAAAC	TTTGGGTGGTGAAGATC	CACAGAGCGACAATTC	1245
Qy	1393	gaatcagtggtcattcaagacagatt	caagcccgacagatagccaact	tttgcggtgaaac	1452
Db	1246	GAGTATGTTTCATTCAAGACCAAT	GATGACCTCGATCGGCACT	TTTGCAAGTGCAAAAC	1305
Qy	1453	tcctcatagataacctgcgcgaggag	gtggttgcaaatcatatggcct	cccaaggag	1512
Db	1306	TCATTGTTGAACGATTCGCGGAG	GAAGTATTACAGAAAAC	TTTTAACCTAAGGAGGCAG	1365
Qy	1513	caggcaagcagcttaagacaacaa	cccccttcaagttctcttcac	acttttttcacgag	1572
Db	1366	CAGGCAGGCGAGGTCAAGAAC	CAACACCCCTTTTCAGTTC	CTCTGGTTCCACCTAAGAGGCT	1425
Qy	1573	ttctccgagggtctggctta	1592		
Db	1426	CAGAGGAGAGATTGTGGCTTA	1445		

```

RESULT 10
US-09-762-381-1
; Sequence 1, Application US/09762381
; GENERAL INFORMATION:
; APPLICANT: Takaiwa, Fumio
; APPLICANT: Utsumi, Shigeru
; APPLICANT: Katsube-Tanaka, Tomoyuki
; TITLE OF INVENTION: TRANSGENIC PLANT EXPRESSING SOYBEAN GLYCININ
; FILE REFERENCE: 201487/1050
; CURRENT APPLICATION NUMBER: US/09/762,381
; CURRENT FILING DATE: 1999-03-04
; PRIOR APPLICATION NUMBER: PCT/JP99/01057
; PRIOR FILING DATE: 1999-03-04
; PRIOR APPLICATION NUMBER: JP 10-223-897
; PRIOR FILING DATE: 1998-08-07
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 1743
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (52)..(1536)
; US-09-762-381-1

```

[illegible]

Db 68 GAGAGCAGCCTCAGCAAAACGAGTGCCAGATCCAAAAACTCAATGCCCTCAAACCGGATA 127


```
Db. 665 GTGGCTTACCCCTGGAATCTTGTGAACATGCAATTCAGCGT---GGACAAGCAGATAGCGA 721
QY 851 aaaatctgtggcgagaaagagagagagagagagagagagagagagagagagagagagagagag 910
Db 722 AAAACCTTACAAGAGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 781
QY 911 gctcagaatcttgagcccc-----gatggaacgagaggtggtccgacgaag 955
Db 782 GTCTGAGCGGTGATAAACCACCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 841
QY 956 aagaggatacagatgaagatcaata-----tgtaataccatgaacagatggaagggc 1006
Db 842 AAGAAGAGAGAGATGAGAAGCCACAGTGCAGGGTAAAGGTAAGACAAACACTGCCAACGCCCCC 901
QY 1007 gtgcaggggaagcagagcgagggggaatgattgaagagacagatcgacacgcagtg 1066
Db 902 GAGAACCCAAAGCAAGCAGAGAAGATGGCAATTGACGAGAGCCATATGACCATTGACATGAGAC 961
QY 1067 taaaaagaacatggtggaacagatccctcacatctacgatccctcagcgctggtta 1126
Db 962 TTGCCACAACATTGGCCAGACTTTCATCACCCTGACATCTACAAACCCTCAAGCCGGTAGCG 1021
QY 1127 ctaaaactgccagatctcaactcttaatacttgaagtggtggtgacttagtgcgtgaat 1186
Db 1022 TCACAACCGCCACAGCCCTTGACTTCCAGCCCTCTCGTGGCTCAGACTCAGTGTGAGT 1081
QY 1187 atgaaatctctacaggaatgctgttctccctcaactacaacacacacacacacacacacac 1246
Db 1082 TTGATCTCTCCGCAAGANTGCAATGTTCTGCCACACTACAAACCCTGAGCGCAACAGCA 1141
QY 1247 tcatatgattgagggcggtccagctgcaagtgggtgacagcaacgcgcaacagag 1306
Db 1142 TAATATACGCAATTGAATGGAGCGGCAATGATACAAAGTGGTGAATTGCACGGGTGAGAGAG 1201
QY 1307 tgtacgagaggagctcaagaggggtcaagcttcttgggtggtgcaagaaactgcgcgtgg 1366
Db 1202 TGTGTGAGAGAGCTGCAAGAGGGAGCGGTGCTGATCTGCCACAAAACCTTGTGGTGG 1261
QY 1367 ctgggaagtcacagagcgaacttcgaatcagctggcattcaagacagagattcaagccca 1426
Db 1262 CTGCAAGATCACAGATGACAACTTCGATGATGTGTCATTCAGACCAATGATACACCCA 1321
QY 1427 gcatagcaactttgocggtgaaactctctcatagataaactgcgcggaggggtggtg 1486
Db 1322 TGATCGGCACTCTGCAAGGGCAAACTCATTTGTTGAACGCATTACCAGAGGAGTGATTC 1381
QY 1487 caattcatatgctcccaagggagcaggaagcagagcttaagaacaacaacaccccttca 1546
Db 1382 AGCACACTTTCAACCTTAAAGAGCCAGAGCCAGGCGAGATGAAGAAACAACACCCCTTCA 1441
QY 1547 agttcttcgttccacacttttcagcagctctccgagggctgtggctta 1592
Db 1442 AGTTCTGGTTCCACCTCAGGAGCTCTCAGNAGAGAGCTGTGGCTTA 1487
```

RESULT 14

```
US-60-312-544-1040
; Sequence 1040, Application US/60312544
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Edgerton, Michael D
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Liu, Jingdong
; APPLICANT: Stein, Joshua
; TITLE OF INVENTION: CDNA SEQUENCES AND USES FOR PLANT IMPROVEMENT
; FILE REFERENCE: 38-10(52726)A
; CURRENT APPLICATION NUMBER: US/60/312,544
; CURRENT FILING DATE: 2001-08-15
; NUMBER OF SEQ ID NOS: 10730
; SEQ ID NO 1040
; LENGTH: 1693
; TYPE: DNA
```

```
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (27)..(1484)
; OTHER INFORMATION: Clone ID: 700676766_FLI
US-60-312-544-1040
```

```
Query Match 27.6%; Score 510.8; DB 64; Length 1693;
Best Local Similarity 61.4%; Pred. NO. 2e-78;
Matches 985; Conservative 0; Mismatches 487; Indels 132; Gaps 5;
```

```
QY 17 acgtttcttttggcttttggcttttggcttttggcttttggcttttggcttttggcttttggctttt 76
Db 34 agcttgcttttcccttttggcttttcccttttccag---tggtgcttctgctctgagagagc 90
QY 77 agccggaggaagatcggtgcaggttccagcgctcctaactgcgcagagacctgacaacccga 136
Db 91 aggcacagcaaaatgagtgccagatcccaaaagctgaatgccctcaaacggataaccgta 150
QY 137 ttgaatcgagggcggtttacattgagacttggaaaccccaacacacagaggttcgaatgcg 196
Db 151 tagagtcggaagtggtttcattgagacctggaaccttaacacaaagccattccagttg 210
QY 197 ccggcgctgcctctctcgtcttagtctccgcgcgaaccccttctgtaggcttttctact 256
Db 211 ccgggtgtgccccctctctcgtcgcaccttaaccgcaatgccccctgtagaccttctaca 270
QY 257 ccaatgtctcccagagagatcttcatccagcaaggaaggatacttttgggttgatatctcc 316
Db 271 ccaacggtcccccagaaatctacatacaacaagtaagtatttttggcatgatattcc 330
QY 317 ctggtgtctcagacacatatgaagcgctgcacaacaagagcgcgatcattcagtcaccaaa 376
Db 331 cgggtgtctcagacttatcaagagcgcgcaaga----- 365
QY 377 gaccaccaagacgtttgcaagaagaagaccaaagcacaagcaacaagacaaagatagtaccaga 436
Db 366 -----tctcagcaacagagacgaagcagagggcccccaagaccgctcaccaaa 411
QY 437 aggtgcacccgtttcaatgaggtgatctcattgcagttcccacccggtgtgtctttcttgc 496
Db 412 aggtacatcgcttcagagaggtgatttgatcgcaagtcctactggtgttcagtggtga 471
QY 497 tgtacaacgacacagacacactgatgtgtgtgtgttctcttactgacaccaacaacaacg 556
Db 472 tgtacaacaatgaagacacactcgtgtgtgtgtgttcttatttatgacaccaacgcttgg 531
QY 557 acaaccagcttgatcagttccccagagatccaatttggctgggaacacacagcaagagt 616
Db 532 agaaccagctcgaccagatgctcagagattctctctgtgtggaaaccaagagcaagagt 591
QY 617 tettaagtlaccagcaacaagacaaaagcagacgaagcagcagaagaagcttaccatatagccat 676
Db 592 ttctaaaaatatacagca----- 610
QY 677 acagcccgcatagtcgctcagcagagagagcgtgaatttcgcccctcgagagacagacaca 736
Db 611 -----gcagcaagaggttctcc 627
QY 737 gcccgagagaacgagcagagacaagaagaagagagagagagagagagagagagagagagagag 796
Db 628 aaagccagaagggaaaagcaacaagaagaagaagaagaagaagaagaagaagaagaagaagaag 687
QY 797 tcacgcccaggttctctggaacaacgcttccaggttgacagacagacagattgtgcacaatc 856
Db 688 tcgccccctgaattcttgaaagaagcggttc---ggcgtgaacatgcagatagtgagaacc 744
QY 857 tgtggggcgagaacgagagtgtaagaagagggagccattgtgacggttgagggagccctca 916
Db 745 tacaaggtgagaacgaagagagagtagtgagccattgtgacagttgaaagagaggtctaa 804
QY 917 gaatttgagcccagatggaacagagaggtgcgcgacgaagaagaggaatacagatgaagatc 976
```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 27, 2001, 13:03:21 ; Search time 8470.79 Seconds
(without alignments)
3608.787 Million cell updates/sec

Title: US-09-715-036-4
Perfect score: 1853
Sequence: 1 atggctaagctcttgtagct.....aaaaaaaaaaaaaaaaaaaaa 1853

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues
Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

- 1: gb.ba.*
- 2: gb.htg.*
- 3: gb.in.*
- 4: gb.om.*
- 5: gb.ov.*
- 6: gb.pat.*
- 7: gb.ph.*
- 8: gb.pl.*
- 9: gb.pr.*
- 10: gb.ro.*
- 11: gb.sts.*
- 12: gb.sy.*
- 13: gb.un.*
- 14: gb.vi.*
- 15: em.ba.*
- 16: em.fun.*
- 17: em.hum.*
- 18: em.in.*
- 19: em.om.*
- 20: em.or.*
- 21: em.ov.*
- 22: em.pat.*
- 23: em.ph.*
- 24: em.pl.*
- 25: em.ro.*
- 26: em.sts.*
- 27: em.sy.*
- 28: em.un.*
- 29: em.vi.*
- 30: em.htgo.hum.*
- 31: em.htgo.inv.*
- 32: em.htgo.rod.*
- 33: em.htg.hum.*
- 34: em.htg.inv.*
- 35: em.htg.rod.*
- 36: em.htg.other.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
------------	-------	-------	--------	----	----	-------------

1	1853	100.0	1853	6	AX148741	AX148741 Sequence
2	1853	100.0	1853	8	AF086821	AF086821 Arachis h
3	1403.4	75.7	1590	8	AF125192	AF125192 Arachis h
4	1391.2	75.1	1524	8	AF093541	AF093541 Arachis h
5	1377.4	74.3	1524	6	AX155334	AX155334 Sequence
6	516	27.8	1743	6	E32414	E32414 Transgenic
7	516	27.8	1743	8	SOYGLYBSU	M36866 Soybean gly
8	512.4	27.7	1708	8	SOYGA2B1A	D00216 Glycine max
9	510.6	27.6	1712	6	E02462	E02462 cDNA encodi
10	510.6	27.6	1712	8	GMGLY1	X02806 Glycine max
11	508	27.4	1746	6	E02463	E02463 cDNA encodi
12	508	27.4	1746	8	GMGLY1A	X02985 Soybean mRN
13	493.4	26.6	1729	8	PSA132614	AJ132614 Pisum sat
14	449.6	24.3	1664	8	VSLEGUMA	Z32835 V.sativa mR
15	439.8	23.7	1666	8	VFLA2	X55014 Vicia faba
16	439.4	23.7	1650	8	VNPROLEGA	Z46803 V.narbonens
17	389.4	21.0	1585	8	VFLA1	X55013 Vicia faba
18	257.6	13.9	4156	8	PSLEGA2	X17193 Pisum sativ
19	255	13.8	3347	8	PSLEGAG	X02982 Pea lega ge
20	242	13.1	787	8	PEALGN	J01255 Pea (P.sati
21	235.8	12.7	1787	8	PATPRU2	X78120 P.amygdalus
22	229.4	12.4	1784	8	GSY5	X79467 G.soja (SH1
23	226.2	12.2	1755	8	AB049440	AB049440 Glycine m
24	223.4	12.1	1763	8	COTSPD	M16905 G.hirsutum
25	223	12.0	1782	8	AB000169	AB000169 Glycine m
26	221.4	11.9	1789	8	AB000168	AB000168 Glycine m
27	208	11.2	3573	8	GMGY3	X15123 Soybean Gy3
28	205.2	11.1	1905	8	PATPRU1	X78119 P.amygdalus
29	204	11.0	1607	8	SOYGLY	K02646 Soybean gly
30	204	11.0	4617	8	GMGY2	X15122 Soybean Gy2
31	202.2	10.9	3176	8	GMGLYCAB	E00868
32	201.8	10.9	1786	6	E00868	E00868 cDNA encodi
33	201.8	10.9	1786	6	E02460	E02460 cDNA encodi
34	201.8	10.9	1786	8	SOYGLYAB	M10962 Soybean (G.
35	201.6	10.9	3527	8	GMGY1	X15121 Soybean Gyl
36	195.2	10.5	1692	8	QRNALP	X99539 Q.robur mRN
37	187.8	10.1	2282	8	CAY15527	Y15527 Cicer ariet
38	187.6	10.1	1674	8	AF319777	AF319777 Glycine m
39	173.4	9.4	1605	8	VSLEGUMB	Z32796 V.sativa mR
40	169.8	9.2	1950	8	GSGLYCGY4	X86970 G.soja mRNa
41	168.2	9.1	1899	6	E02461	E02461 cDNA encodi
42	168.2	9.1	1899	8	GMGLYR1	X02626 Soybean mRN
43	162.6	8.8	1939	8	SSPLEGUM	Y09116 S.sagittifo
44	160.8	8.7	2826	8	HNNHAG3D1S	M28832 Sunflower 1
45	159.4	8.6	1679	8	MSLEGA11	X82463 M.salicifol

ALIGNMENTS

RESULT	1
LOCUS	AX148741 1853 bp DNA
DEFINITION	Sequence 4 from Patent WO0136621.
ACCESSION	AX148741
VERSION	AX148741.1 GI:14347295
KEYWORDS	peanut.
SOURCE	Arachis hypogaea
ORGANISM	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Aeschynomeneae; Arachis.

REFERENCE	1 (bases 1 to 1853)
AUTHORS	Dodo,H.W.; Arntzen,C.J.; Konan,K.N. and Viquez,O.M.
TITLE	Down-regulation and silencing of allergen genes in transgenic peanut seeds
JOURNAL	Patent: WO 0136621-A 4 25-MAY-2001;
FEATURES	Alabama A & M University (US) Location/Qualifiers

```

source      1..1853
            /organism="Arachis hypogaea"
            /db_xref="taxon:3818"
BASE COUNT  562 a 430 c 465 g 396 t
ORIGIN
Query Match 100.0%; Score 1853; DB 6; Length 1853;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1853; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 atgcataagctcttgagcttctctttgttgctttgtcttcttagttcttgaggagctagcagc 60
    |||||
Db 1 ATGCCTAAGCTTCTTGAGCTTTCCTTTTGTCTTTGCTTTCTAGTCTTGGGAGTAGCAGC 60
    |||||

Qy 61 atctcttcaggcagcagcggaggagaaatcgtygccagttccagcgctcaatgcgcag 120
    |||||
Db 61 ATCTCTTCAGGCAGCAGCGGAGGAGAAATCGTCCAGTTCAGCGCCCTCAATGCGCAG 120
    |||||

Qy 121 agacctgacaacgcattgaatcggaggcggtttacattagaccttgaaaccccaaac 180
    |||||
Db 121 AGACCTGACAACCGCATTAATCGGAGGGCGGTTACATTGAGACTTGGAAACCCCAACAC 180
    |||||

Qy 181 caggagttcgaaatgcgcggcgctgcctctctctcgcttagtctctccgcgcgaacgcctt 240
    |||||
Db 181 CAGGAGTTCGAATGCGCGCGCTCGCCCTCTCTGCTTAGTCTCCGCCCAACGCGCTT 240
    |||||

Qy 241 cgtaggcctttacttccaatgctccccaggagatcttctatccagaagaaggagatcac 300
    |||||
Db 241 CGTAGGCGCTTTACTTCCAATGCTCCCCAGGAGATTTTCATCCAGCAAGAAAGGGGATAC 300
    |||||

Qy 301 ttggggttgatatccctgggtgtctctagcacatatgaagagcctgcacacaagagcgc 360
    |||||
Db 301 TTTGGGTTGATATTCCCTGTTGCTCTAGCACATATGAAGAGCTGCACAAACAAGGACGC 360
    |||||

Qy 361 cgatatcagtcctcaagagacacacaagacgcttgcagaagaagaccacaaacagagcaa 420
    |||||
Db 361 CGATATCAGTCTCCAAAGACACCAAGAGCTTTGCAAGAAGAAGAACCAAGGCCAACAGCAA 420
    |||||

Qy 421 caagatagtcaccagaggtgtcacgctttccaatgagggtgatctcattgcagttcccaacc 480
    |||||
Db 421 CAAGATAGTCAACAGAGGTGCACCGCTTTCATGAGGGTGATCTCAITGCAGTTCGCCACC 480
    |||||

Qy 481 ggtgtgtctttcgtgtgtatcacacgaccacgacctgattgtgtgtgttctcttact 540
    |||||
Db 481 GGTGTGCTTCTTGCGTGTACAACGACCACGACACTGATGTTGTGTGTCTTCTTACT 540
    |||||

Qy 541 gacccaacaacaacacacacacagcttgatcagttcccccaggagatcaatttggctggg 600
    |||||
Db 541 GACACC AACACACACACACACACAGCTTGATCAGTTGCCCCAGGAGANTCAATTGGCTGGG 600
    |||||

Qy 601 aaccacgagcaaggttctttaagttaccagcacaaagcagacaaagcagacgaagaagc 660
    |||||
Db 601 AACCACGAGCAAGTTCCTTAAGGTACCAGCAACAAGCAGACAAAGCAGACGAAGAAGC 660
    |||||

Qy 661 ttacatatagcccatatcagcccgcatagtcggcctagacgagaagagcgtgaatttcgc 720
    |||||
Db 661 TTACCATATAGCCCATACAGCCCGCATATGTCGCCCTAGCAGAGAAGAGCGTGAATTTCCG 720
    |||||

Qy 721 cctcgagacagcacagcccgacagaaacgagcagggaacgaagaagaagacgaaggtgga 780
    |||||
Db 721 CCTCGAGGACGACACGCCGACAGAACGACGAGCAAGAAAGAAAGAGACGAAGGTGGA 780
    |||||

Qy 781 aacatcttcagcggcttccacgcgcggagttctctggaacagccttcaggttgacacaga 840
    |||||
Db 781 AACATCTTTCAGCGCTTCACGCGCGAGTTCTCTGGAAACAAGCCTTCCAGGTTGACGACAGA 840
    |||||

Qy 841 cagattgtcaaatctgtgggcgcgaacgagagtggaagagagggagcctatgtgacg 900
    |||||
Db 841 CAGATTGTGCTAAAATCTGTGGGCGGAGACGAGAGTGAAGAAGAGGGAGCCATTGTGACG 900
    |||||

Qy 901 gtgaggggagccttcagaaactttgagcccgga*ggaacgagaggtgcgcgacgaagaag 960
    |||||

```


Db 542 AAGAGTTCTTAAGGTACCAGCAACAAGACAGACAAGACGACGAAGAGCTTACCATATA 601
Qy 671 gccatacagccgcctagctcgccttagacgagaagacgctgaattcgcctcgcagac 730
Db 602 GCCCATACAGCCGCCAAGTCAGCTTAGACACAGAGAGGCTGAATTAGCCCTCAGGAC 661
Qy 731 agcacagccgcagagacgagcaggacaagaagaagaagcgaagcgaagcgaacatcttca 790
Db 662 AGCACAGCCGACAGAGACGAGGACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 721
Qy 791 ggcgttcacgcgcggaggttcttgcgaacagccttcagggttacacagacagagattgac 850
Db 722 GCGGCTTCAGCGCGAGTTCCTGGAAACAAGCCTTCAGGTTGACGACACAGACAGATAGTC 781
Qy 851 aaaaatctgtgggcgagaagcagagatgaagaagagagagccattgtgacggtgagggag 910
Db 782 AANAACCTAGAGGCGAGAGCGAGAGTGAAGAGAGGAGGCCATTGTGACAGTGAAGGGAG 841
Qy 911 gctcagaatcttgagccagatggaaacgagaggtgcccacacgaagaagagggaatcacgtg 970
Db 842 GCCTCAGAATCTTGAGCCCGATAGAAAGAGAGCGTCCGACGAGAGAGAGAGATACGATG 901
Qy 971 aagatcaatgaataaccatgaacaggaatggaagcgctggcggggaagcagagcgggg 1030
Db 902 AAGATGAATATGAATACGATGAAGAGAGATAGAAGGCGTGGCAGGGGAGACGAGAGCGG 961
Qy 1031 ggaatgattgaagacagatcgcacccatgtttaaagaagaacattggtgaacaa 1090
Db 962 GGAATGGTATTGAAGAGAGCATCTGCACCGCAAGTCTAAAAGAAACATTTGGTAGAACA 1021
Qy 1091 gatccctcacatctacgatccctcagcgctggttcaactcaaaactgcc-acgatctcaac 1149
Db 1022 GATCCCTGACATCTACAACCTCA-AGCTGGTTCACTCAAAACTGCCAACGATCTCAAC 1080
Qy 1150 cttctaatccttgggttgacttagtgcgaataatggaatactctcaaggaatgca 1209
Db 1081 CTCTTAATACTTGGTGGCTTGGACCTAGTGTGCTGAATATGGAATCTCTACAGGATGCA 1140
Qy 1210 ttgttgcctcactacaac 1269
Db 1141 TTGTTTGTGCTCAGTACAAC 1200
Qy 1270 gctcagctgcaagtgtggacagcaacgcaacagagtgctacgacgagagcttcaagag 1329
Db 1201 GCTCAGCTGCAAGTGTGGACACCAACGCAACAGAGTGTACGACGAGGAGCTTCAAGAG 1260
Qy 1330 ggtcaagcttcttggctgcacagaacttcgcgctggctggcgaagtcacagagcgaac 1389
Db 1261 GGTCACTGCTTGTGTGTCACAGAACTTCGCGCTGCTGGAAAGTCCAGAGCGAGAAC 1320
Qy 1390 ttcgaatacgtggcattcaagacagacattcaagccacagacacacacacacacacacac 1449
Db 1321 TTCGAATACGTGGCATTTCAAGACAGACTCAAGGCCACAGATACCCAACTCGCGGTGAA 1380
Qy 1450 aactcctctatagataacctgcgcggagggtgtgttgcaaaatcatatgacctcccaagg 1509
Db 1381 AACTCCGTCATAGTAACCTGCCGAGGAGTGTGTGCAAAATTCATATGGCCTCCAAAG 1440
Qy 1510 gacgaggaaggcagcttaagaacaacacacacacacacacacacacacacacacacacac 1569
Db 1441 GAGCAGGCAAGGAGCTTAAGAACAACAACACCCCTTCAAGTTCTTCGCTCCACGCTCAG 1500
Qy 1570 cagtctccgagggctgtgacctaa 1593
Db 1501 CAGTCTCCGAGGGCTGTGGCTTAA 1524

RESULT 5
AX155334
LOCUS AX155334 1524 bp DNA PAT 22-JUN-2001
DEFINITION Sequence 7 from Patent WO0140264.
ACCESSION AX155334
VERSION AX155334.1 GI:14536769

KEYWORDS peanut.
SOURCE Arachis hypogaea
ORGANISM Arachis hypogaea
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae;
Aeschynomeneae; Arachis.
REFERENCE 1 (bases 1 to 1524)
AUTHORS Bannon,G.A., Burks,W.A., Caplan,M.J., Sampson,H. and Sosin,H.
TITLE Peptide antigens
JOURNAL Patent: WO 0140264-A 7 07-JUN-2001;
Panacea Pharmaceuticals, LLC (US); The University of Arkansas (US)
; MOUNT SINAI SCHOOL OF MEDICINE OF THE CITY UNIVERSITY OF NEW YORK (US)
FEATURES
source 1..1524
/organism="Arachis hypogaea"
/db_xref="taxon:3818"
BASE COUNT 456 a 392 c 395 g 281 t
ORIGIN
Query Match 74.3%; Score 1377.4; DB 6; Length 1524;
Best Local Similarity 94.9%; Pred. No. 0;
Matches 1445; Conservative 0; Mismatches 76; Indels 2; Gaps 2;
Qy 72 gcagcagccgagagagaatcgctgccagttccagcgcctcaatcgcgagagacctgacaa 131
Db 3 GCAGCAACCGGAGGAGACGCGTCCAGTTCACAGCGCTCAATCGCAGAGACCTGACAA 62
Qy 132 cgcattgaatcgagggcggttacattgagacctggaaccccccaacacaggaggtcga 191
Db 63 TCCCATTTAATCAGAGGCGCGTTTACATTGAGACTTGGAAACCCCAACACAGGAGTTCCA 122
Qy 192 atgcgcggcgctgcctctctcgttagtctcgcgcgcgcgcgcgcgcgcgcgcgccttt 251
Db 123 ATGCGCGCGCTCGCCCTCTCTCGCTTAGTCTCTCGCGGAAACGCCCTTCGTAGGCCCTT 182
Qy 252 ctactccaatgctcccagagatcttcacccaaggaaggggatactttgggttgat 311
Db 193 CTACTCCAATGCTCCCCAGAGATCTTCATCCAGCANAGGAGGATACTTTGGGTTGAT 242
Qy 312 attcctggttgcttagcacatatgaagagcctgcacacacacacacacacacacacacacac 371
Db 243 ATTCCGTGTTGTCTTAGACACTATGAAGAGCCTCACACACAAGGTGCTCGATCTCAGTC 302
Qy 372 ccaagacacacacagacgtttgcaagaagaagacacacacacacacacacacacacacac 431
Db 303 CCAAGACACCAAGACGTCTCCAAGGAGAGACCAAGCCCAACAGCAACAGAGATAGTCA 362
Qy 432 ccagaggtgcacgcgtttcaatgaggtgagctcctcagttcccccacgcggtgtgcttt 491
Db 363 CCAGAAAGTGCACCGTTTCGATGAGGTTGATCTCATTTGCAGTTCCACCGGTGTGCTTT 422
Qy 492 ctggcgtgtacaac 551
Db 423 CTGGCTCTACAACGACACACACACTGATGTTGTTGCTGTTTCTTCTTACTGACACACAA 482
Qy 552 caacgac 611
Db 483 CAACGACAAACAGCTTGATGATTTCCCGCAGGAGATTCAATTTGGCTGGGAACACGAGAC 542
Qy 612 agagttcttaaggtaccagcaacaaagcagacaaagcagacgaagaagccttaccatag 671
Db 543 AGAGTTCTTAAAGTTACAGACCAACAAAGCAGACAAAGCAGACCAAGAGCTAACCATATAG 602
Qy 672 ccatacagccgcctagtcgcgccttagacgagagcgtgaatttcgcctcgcagaca 731
Db 603 CCCATACAGCCGCAAGTCAAGCTAGACCTTAGAAGAGAGAGCGTGAATTTAGCCCTCGAGGACA 662
Qy 732 gcacagccgcagac 791
Db 663 GCACAGCCGCGAGAGAGGAGGACCAAGAGAGAGAAACCAAGAGTGGAAACATCTTCAG 722

Db	645	-----		-----GCAAGGAGGTC	655
QY	731	agcacagccgcagagaacacgacgagcaagaagaagaagacgaagtggaacacattcca	790		
Db	656	ATCAMAAGCCAGAAAGGAAGCATCAGCAAGAAGAAGAAACGAAGGAGGAGCAGATATTGA	715		
QY	791	gcggttcaacgcgagttcctggaacaagccttccagggttgacacacagacagattgtgc	850		
Db	716	GTGGCTTCAACCTGGAAATCTTTGGAACATGCAATTCACGGT---GGACAACGAGATAGCGA	772		
QY	851	aaattctgtggtgcgagaaacagagatgaagaagaggagcattgtgacggtgaggggag	910		
Db	773	AAANCTTACAAGAGAGAACGAAGGGAAGACAAGGGACCATTTGTGACAGTGAAGGAG	832		
QY	911	gcctcagaatcttgagccca-----gatgaaacagaggtgcccacgaagaag	955		
Db	833	GTCTGAGCGTGATAAACCACCCACGAGCAGCAGCAACAAGACCCCGAAGAAGGAAG	892		
QY	956	aagagaatactgatgaagatacaata-----tgaataccattgaacagatggaaggc	1006		
Db	893	ANGAAGAAGAGATGGAAGCCACAGTGCACAGGGTAAAGACAAACACTGCCACGCCCC	952		
QY	1007	gtgcgagggaagacagagcggggggaatggtattgaagacagatctgcaccgcgatgtg	1066		
Db	953	GAGGAGCCCAAGCAAAAGCACAAGAANTGCAATTGACGAGCCATATGCACCATGAGAC	1012		
QY	1067	ttaaaagaacatttggtggaacagatccctcacaatctacgactcttcacgcgtgtgttca	1126		
Db	1013	TTCCGCACAACATTGGCCAGACTTCATCCCTGCATCTACAACCCCTCAACCGGTAGCG	1072		
QY	1127	ctcaaaactgcacgatctcaacaccttcaactcttagtggcttggacttagtgcctgaat	1186		
Db	1073	TCACAACCGCCACAGCCCTTGACTTCCGAGCCCTCTCGTGGCTCAGACTCAGTGTCTGAGT	1132		
QY	1187	atggaatctctacaggaatgatgtgttgcctcacaacacacacacgacacagca	1246		
Db	1133	TTGGATCTCTCCGCAAGANTGCAATGTTCTGTCGCCACACTACAACCTGAACGCGAAGCA	1192		
QY	1247	tcatatatgattgagggagcgggctcactgctgaagtgtggacagacaacggcaacagag	1306		
Db	1193	TAATATACGCAATTGAATGGACGGGCAATTCATACAAGTGGTGAATGTCAACCGGTGAGAG	1252		
QY	1307	tgtacacagagacttcaagagggtccagcttcttgggtgccacgaacttcgcgttg	1366		
Db	1253	TGTTTGATGAGAGCTGCAAGAGGGACGGTGCTGATCGTGCACCAAACTTTGTGTGTG	1312		
QY	1367	ctgggaagtccccagagcgaacttcgaatcagctggcattcaagacagattcaaggccca	1426		
Db	1313	CTGCAAGATCACAGATGACAACTTCGAGTATGTGTCATTTCAAGACCAATGATACACCCA	1372		
QY	1427	gcatagcacatttgcggtgaaaaactcctcatagataaactgccggagagagtgttg	1486		
Db	1373	TGATCGCACCTCTTGACGGGGCAACTCAATTTGTAACGCATTTACGAGAGAGTGAATC	1432		
QY	1487	caaatctatagccttcccaaggagcagcagcagcagcagcagcagcagcagcagcagcagc	1546		
Db	1433	AGCAACTTTCACCTTAAAGCCACGACGCGCAGCAGATTAAGACACAACACCTTTCA	1492		
QY	1547	agttcttcgttccaccttttccagcagcttcccgagggtctgtggcttaaa	1594		
Db	1493	AGTTCTCTGGTTCCACCTCAGGAGTCTCAGAAGAGAGCTGTGCTTAGA	1540		
RESULT	7				
SOYGLYBSU					
LOCUS	SOYGLYBSU	1743 bp	mrna	PLN	27-APR-1993
DEFINITION	Soybean glycinin A-1a-B-x subunit	mrna, complete cds.			
ACCESSION	M36886	D00566			
VERSION	M36886.1	GI:169972			
KEYWORDS	glycinin; storage protein.				
SOURCE	Soybean (var. Shirotsurunoko)	cotyledon, cDNA to mRNA, clones pGST18,4-2-3,4-2-11-101.			

ORGANISM Glycine max
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.

REFERENCE 1 (bases 1 to 1743)
AUTHORS Utsumi,S., Kohno,M., Mori,T. and Kito,M.
TITLE An alternate cDNA encoding glycinein A-la-B-x subunit
JOURNAL J. Agric. Food Chem. 35, 210-214 (1987)
FEATURES Location/Qualifiers
source 1..1743
 1 (bases 1 to 1743)
 /organism="Glycine max"
 /db_xref="taxon:3847"
 /tissue_type="cotyledon"
 /feature_id="AAA33966.1"
 /translation="MAKLVFLSCFLFSGCCFAFSRRSPQOONCOIQKLNALPKDNRN
 TESGGLIETWNNPFCQAGVALSRHCLTNALRRSPYNGPQEIYIOGKGIQGV
 IYPCPTFFEEPQOQRQSPQRKQIYNFREGDLIAVPTGVAWMYNNEETPV
 VAVSIIDTNLQDPMPRFYLAGNQEDELKYQQGGVGSQKHQGEENEGG
 SILGFLTELEHAFSDFQIAKNLQGENEGEKGAIVTVKGLSVIKPPYDEQQQRP
 EEEEEEDTPOCKKDKHCKQRPGRSQSKRRNGIDETICTMLRLHNIIGTSSPDYI
 NQPAZSTATSDIFDPALSWLRLSAEFSLRKNAMFPHYNLANSIYIALNGRALIQ
 VYVNGSRVFDGELGSRVLIYVQNFVVAARSQSNFVSKTNDTPTMCTLAGANS
 LNLALPEVIOHTFNLKSQQARQIKNNPFKFLVLPQESQRAVA"
 polysignal 1597..1602
 polysignal 1711..1716
BASE COUNT 552 a 419 c 398 g 374 t
ORIGIN

Query Watch 27.8%; Score 516; DB 8; Length 1743;
Best Local Similarity 62.3%; Pred. No. 5.1e-114;
Matches 1002; Conservative 0; Mismatches 450; Indels 156; Gaps 6

Qy 17 agctttcttttctgtcttcttcttcttctgtcttctggagctagcagcatctctt-----ca 70
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 59 AGCTAGTTTTTCCCTTGTCTTCTGCTTTTCAGTGGCTGCTGCTTCAGTTCCA 118
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 71 ggcagcagccggagaggaatgcgtccaggttcgcgcctcaatgcagcagacctgaca 130
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 119 GAGAGCAGCCTCAGCAAAACGAGTCCAGATCCAAAACACTCAATGCCTCAACCGGATA 178
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 131 accgattgaatcgagagcggttaccattgagacttggaaccccaacacgagaggtctg 190
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 179 ACCGTATAGAGTCAGAAGAGGGCTCATTTGAGACATGGAACCTTAACAACGCCATTCC 238
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 191 aatgcgcgcggctgcgcctctctctcttagtcttcgcgcgcgaacgcctctcgtagccct 250
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 239 AGTGTGCGGTGTGTGCCCTCTCTCGCTGCAGCCCTCAACCGCAACGCCCTTCGTAGACCTT 298
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 251 tctactcaaatgctcccaagagatcttcattccagcagaaggggtactcttgggttga 310
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 299 CCTACCAACACGGTCCCCAGGAATCTACATCCAACAAAGTTAAGGTTATTTTGGCATGA 358
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 311 tattccctgggttctctagcacatatgaagagcctgcacaacaaggacgcgcatatcagt 370
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 359 TATACCGGGTTGTCTAGCACATTTGAAGAGCCTCAACAAC----- 400
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 371 cccaaagaccacaagaagtttgaagaagaagaaccaagccaacgaacgaacgaacgaatgc 430
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 401 -----CTCAACAAGAGGACAAAGCAGCAGCAGACACACAGACCGCTC 439
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 431 accgaaggtgcacggttttcaatgaggggtgatctcattgcaggttccaccacgggttgctt 490
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 440 ACCAAGATCTATAACTTCAGAGAGGGTGATTTTGATCGCAGTGCTACTGTTGTCAT 499
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 491 tctggctgtacaacgaccacgacactgattgttgttgtcttcttcttactgacacaaca 550
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 500 GGTGGATGTACAACATGAAGACACTCCCTGTGTGTGGCTCTTCTATTATTGACACAACA 559
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```
QY 551 acaacgacacacagcttgatcagttccagcagagattcaatttggctggaaacacagc 610
D 560 GCTTGGAGAACCCAGCTGCACCATGCTAGAGATTCTATCTTGTGCGAACCAAGAG 619
QY 611 aagagttcttaagtaccagcaacaagcagacaaagcagacgaagaagcttaccatata 670
D 620 AAGAGTTTCTAAAATATCAGCAAGA----- 644
QY 671 gccatatacccgcatagtcggtccttagacagagaagcgtgaattcgccttcgagagac 730
D 645 -----GCAAGGAGGTC 655
QY 731 agcacagccgcagagaacagcagcagacaagaagaagcgaagtggaactcttca 750
D 656 ATCAAGCCAGAAAGGAAGCATCAGCAAGAGAAAGAAAGCAAGAGGAGGAGCATATTGA 715
QY 791 gcgcttcacgcgcaggttctctggaacaaagccttcaggttgacgacagacagattgtgc 850
D 716 GTGGCTTTCACCTCGAATCTTGAACATGATTCAGCGT---GGACAAGCAGATAGCGA 772
QY 851 aaaatcgtggtgagcagcagagagtaagaagaagagagcattgtgacggtgagggag 910
D 773 AAAACCTACAGGAGAGAGCAAGAGGGGAAGAGGAGGAGCCATTGTACACAGTGAAGGAG 832
QY 911 gctcagaatcttgagccca-----gatggaacgagaggtgcccagcgaag 955
D 833 GTCTGAGCGTGATAAAACCCAGCCAGCAGCAGCAACAAACCCAGAGAGAGGAAG 892
QY 956 aagaggaaatcagatgaagatacaata-----tgaataccatgaacaggtggaagcc 1006
D 893 AAGAAGAGAGGATGAGAAGCCACAGTGCAGGGTAAAGCAAAACACTGCCAACGCCGCC 952
QY 1007 gtggcaggggaagcagagcgcggggaatgtattgaagacagatctgcacgcgatgtg 1066
D 953 GAGAAACCCAAAGCAAGAAAGAAATGGCAATGACGAGACCATATGCACCATGAGAC 1012
QY 1067 ttaaaagaacattggtggaacacagatccctccatctacagctctccgcgctgttca 1126
D 1013 TTGCCACAACATTTGCCAGACTTCATCACCTGACATCTACAACTTCAAGCCGTAGCG 1072
QY 1127 ctcaaaactgcacagatctcaactcttaactcttggtggtgacttagtctgtaat 1186
D 1073 TCACACCGCCACACAGCTTGACTTCCAGCCCTCTCGTGGCTCAGACTCAGTGTGAGT 1132
QY 1187 atggaatctctacaggaatgcttctgctccactacacacacacacacacagca 1246
D 1133 TTGGATCTCTCCGACAGATGCAATGTTGTGCCACACTACAACTGAACCGGAACAGCA 1192
QY 1247 tcatatatgcatgagggagcggtcacgtgcaagtgtggaacagcaacgcgcaacagag 1306
D 1193 TAATATACGCATTGAATGGAGCGCATTTGATACAAGTGGTGAATTCACGGTGAAGAG 1252
QY 1307 tgtacagcagagagcttcaagaggggtcaagcttctgtgtggtgcaacaaactgcgctgg 1366
D 1253 TGTTGTGAGAGAGCTGCAAGAGGGAGCGGTGTGTGCTGTCGCACAAAACCTTTGTGGTG 1312
QY 1367 ctgggaagtcacagcagcagaaactcgaatacgtggtggtggtggtggtggtggtggtg 1426
D 1313 CTGACAGATCAGAGTGCACAACTTCGATGTGTGTCATTCAGAGCAATGATACACCA 1372
QY 1427 gcataccaacttgcggtgaaactccttcatagataaacctgccgagggaggtggtg 1486
D 1373 TGATCGGCACTCTGCAGGGGCAAACTCATTTGTTGACCCATTACAGAGAGGAGTATC 1432
QY 1487 caaattcatatggtcctcccaagggagcagcagcaggttaagaacaacaaccccttca 1546
D 1433 AGCACACTTTCAACCTTAAACCTTAAAGAGCAGAGCCAGGAGGAGGAGGAGGAGGAGG 1492
QY 1547 agttcttccttcacaccttccagcagctcccgagggctgtgctttaa 1594
D 1493 AGTTCCTGGTTCCACCTCAGGAGTCTCAGAGAGAGAGCTGTGGCTTAGA 1540
```

RESULT 8

```
SOYGA2B1A 1708 bp mRNA PLN 09-JAN-1998
LOCUS Glycine max mRNA for glycinin A2B1a subunit, complete cds.
DEFINITION D00216
ACCESSION D00216
VERSION D00216.1 GI:218264
KEYWORDS A2B1a; glycinin subunits.
SOURCE Glycine max (variety:Shiotsurunoko) mid-maturation stage cotyledon
        cDNA to mRNA, clone:pgST4-3-1-4.
ORGANISM Glycine max
        Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
        Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots;
        Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
        Glycine.
REFERENCE 1 (bases 1 to 1708)
AUTHORS Utsumi,S., Kim,C.-, Kohno,M. and Kito,M.
TITLE Polymorphism and expression of cDNAs encoding glycinin subunits
JOURNAL Agric. Biol. Chem. 51, 3267-3273 (1987)
COMMENT The nucleotide sequence of cDNA encoding the glycinin A2B1a subunit
        from var. Shiotsurunoko was determined and compared with those of
        var. Bonminorori and CX635-1-1-1.
        Location/Qualifiers
            1..1708
                /organism="Glycine max"
                /variety="Shiotsurunoko"
                /db_xref="taxon:3847"
                /clone="pgST4-3-1-4"
                /tissue_type="cotyledon"
                /dev_stage="mid-maturation stage"
                /note="48 bp upstream of Hind III site"
            42..1499
                /codon_start=1
                /product="glycinin A2B1a subunit"
                /protein_id="BAA00154.1"
                /db_xref="GI:218265"
                /translation="MAKLVLSCFLFLFSCGFALPEQAQNECOIQKLNALKPDNRITES
                EGGFLEWPNPNKPFQCGAVALSCTLNRLALRPSTNGPQEIYIQGGNGIFGMIFP
                GCPSTYQPSQQRGRSQRPDRHQVHRFREGDLIAVTPGVAMWYNNEDTFVAV
                SIIDTNSLENDQMPRRFLVLAGNQEFELKYQQQGGSGSQKQKEEGSNIL
                SGFAPEFLKEAFVGMQIVRNLOGENEEDSGAIVTVKGLRVTPAMRRKPOGEDDD
                DEEPQPCVETDKGCRQSKRSNGIDETICTMRLRONIGONSPLDYNPQASITTA
                TSLDFPALWLLLSAOKYSLRKNAMFPHYTLNANSIYALNGRALQVYVNCGERVF
                DGELEQGVLLVLPQNEFAVAAKQSDNEFYVFTNDPRPSIGNLAGANSLNALPEEVI
                QHTNLSQARQKNNPFSFLVPQESORRAVA"
            95
                /note="t in var. Shiotsurunoko changed to c in var.
                Bonminorori"
                /replace="c"
            96
                /note="c in var. Shiotsurunoko changed to t in var.
                Bonminorori"
                /replace="t"
            155
                /note="g in var. Shiotsurunoko changed to a in var.
                Bonminorori"
                /replace="a"
            157
                /note="a in var. Shiotsurunoko changed to g in var.
                Bonminorori"
                /replace="g"
            1265
                /note="c in var. Shiotsurunoko changed to t in var.
                Bonminorori and CX635-1-1-1"
                /replace="t"
            1460
                /note="g in var. Shiotsurunoko changed to t in var.
                Bonminorori, and to a in var. CX635-1-1-1"
                /replace="a"
            1511
                /note="t in var. Shiotsurunoko deleted in var. Bonminorori"
                /replace=""
```

variation	1533	/note="t in var. Shiotsurunoko deleted in var. Bonminori"	
		/replace=""	
polya_signal	1577..1582	/note="putative"	
variation	1616		
		/note="t in var. shiotsurunoko changed to c in var. Bonminori"	
		/replace="t"	
polya_signal	1688..1693		
		/note="putative"	
BASE COUNT	519 a 399 c 414 g 376 t		
ORIGIN			
Query Match	27.7%;	Score 512.4;	DB 8; Length 1708;
Best Local Similarity	61.5%;	Pred. No. 3.8e-113;	
Matches	986; Conservative	0; Mismatches 486; Indels 132; Gaps 5;	
Qy	17	agcttcttttggcttcttggcttctgagctagcagcatctcttcaggcagc	76
Db	49	AGCTTGTCTTTCCCTTTGTTTCCCTCTTTTCAG---TGGCTGCTTGGCTGTGAGAGAC	105
Qy	77	agccgagagagaatgcgtgcagttccagcgcctcaatgcgcagagacctgcaaacccga	136
Db	106	AGCAGAGCAAAATGAGTCCAGATCCAAAGACTGAATGCCCTCAAACGGATAACCGTA	165
Qy	137	ttgaatcgaggcgggtttacattgagactttggaaccccaaacccaggagttcgaaatgcg	196
Db	166	TAGATCGGAAGGTGGGTTTCATTGTAGACTCGGAACCCCTAAACAAGCCATTCCAGTGTG	225
Qy	197	ccggctgcctctctgcgttagtcttcgcgcgcgaacgccttcgttagcctttctact	256
Db	226	CCGGTGTGCCCTCTCTCGCTCGACCCCTTAACCGCAATGCCCTTCGTAGACCTTCCTACA	285
Qy	257	ccaatgtcccccaggagatcttcatccagcaaggaaggagataccttgggtttgatattcc	316
Db	286	CCAACGGTCCCCAGGAATCTACATACAAAGGTAATGGTATTTTTGGCATGATATGCC	345
Qy	317	ctggttgtctagcacatatgaagacctgcacaaaggacgcgcatatcagtcaccaaa	376
Db	346	CGGTTGTCTTAGCACTTATCAAGAGCCGCAAGAA-----	380
Qy	377	gaccaccaagacgtttgcaagaagaagaccaaagccaacagcaacaagatagtcaccaga	436
Db	381	-----TCTCAGACACGAGGAGCAAGCAGAGGCCCAAGACCGTCACCAAA	426
Qy	437	aggtgcaccttcaatgaggggtgattctcaatgcaagttcccaacgggtgtgtctttctggc	496
Db	427	AGGTACATCTGCTTCAGAGAGGGTGATTTGATCGCAGTGCCTACTGGTGTTCATGTGGTGGGA	486
Qy	497	tgtacaacgaccacgacctgattgttgtcttcttactgcaccacaaacaaacg	556
Db	487	TGTACAAATGAAGACATCTCTGTTGTGCCGTTCTATTATTGACACCAACAGCTTGG	546
Qy	557	acaaccagcttgatcagttcccccaggagattcaatttggcttgggaaccacgacgaaggt	616
Db	547	AGAACCAAGCTCGACCAGATGCCCTAGGAGATTCTATTCTGCTGGGAACCAAGCAAGAGT	606
Qy	617	tcttaaggtaccagcaacaagacagacaaagcagcagaagaagcttaccatatagcccat	676
Db	607	TCTTAAATAATCAGCAGCA-----	625
Qy	677	acagccgcgatagtcggcctagacgagaagacgtgaatttcgccttcgaggcagcagcaca	736
Db	626	-----GCAGCAAGGAGGTTTCCC	642
Qy	737	gccgcagagaaacgagcagcagaagaagaagacgaagggtggaaaaaatcttcaaggcgt	796
Db	643	AAAGCCAGAAAGGAAACCAACAGAAAGAAAGAAACCAAGGAAGCAACATATTTAGTGGCT	702
Qy	797	tcacgcgaggttccctggaaacagccttcaggtttgacgacagacagatttgtcaaaatc	856

QY	1367	ctgggaagtcacagagagaaacttcgaataacatcgatcgacattcaagacagattcaaggccca	1426
DB	1316	CTGCAAGATCATCAGAGTGCACAACTTCGAGTAGTGTGCTCATTCAGAGCAATATACACCCA	1375
QY	1427	gcataagcaacttgcggtgaaactccttcataataaactccgcgagggaggtggtg	1486
DB	1376	TGATCGGCACACTTCGACGGGCAACATCATGTTTGACGCATTACACAGAGGAATGATTC	1435
QY	1487	caaatcatatgcttcccaaggagcagcgagcgagcttaagaaacaacaccccttca	1546
DB	1436	AGCACATTTCAACCTAAAAGCCAGCAGCGCAGGACAGATAAAGAACAAACAACCCCTTCA	1495
QY	1547	agttcttgctccacttttcagcagtcgtccgagggctgtggtctaaa	1594
DB	1496	AGTTCCTGTTCCACCTCAGGAGTCTCAGAAGAGAGCTGTGGCTTAGA	1543

RESULT

12

GMGLYL1A

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

1746 bp

mRNA

PLN

21-MAR-1995

Soybean mRNA for glycinin AlaBx precursor.

X02985

X02985.1

GI:18614

glycinin; seed storage protein; signal peptide.

soybean.

Glycine max

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

COMMENT

The precursor is post-translational processed to form a covalently linked Ala-Bx subunit complex.

1. (bases 1 to 1746)

Negoro,T., Momma,T. and Fukazawa,C.

A cDNA clone encoding a glycinin Ala subunit precursor of soybean Nucleic Acids Res. 13 (18), 6719-6731 (1985)

86041867

FEATURES

source

sig_peptide

CDS

Location/Qualifiers

1..1746

/organism="Glycine max"

/db_xref="taxon:3847"

55..111

/note="putative"

55..1542

/note="glycinin precursor"

/codon_start=1

/protein_id="CAA26723.1"

/db_xref="GI:18615"

/db_xref="SWISS-PROT:P04776"

/translation="MAKLVFSLCLFLSGCCFAFSREPOQNECQIQKLNALKPGNHNISEGGLLETWNPNNPQCAVALSRCTLNRLNRRPTNGPQETIYQKGKIFGNITYPGCSFTFEPQDPQGGSSRPDRHQKIYNSREGDLIAVPTGVAMWYNNEDTPVAVSIIIDTNSLENDLQDMPRRFYLAGNQEQEFLKYQEGGQSGKQKHQEEQENEGSILGFTLEFLHAFSVDKOIAKNLOGENEGDKGAIIVTKGSLSVIKPPTDEQQQRIFSEEEDEEPCQCKKHCORPGSSRRNGIDETICTMRLRHNIIGOTSSPDIDNQAGSVTATSLDFPALSRLWLSLACGSLRKNAFVHNLNANSILVALNGRALIQVNLNPERVFQEGEGRGLVLPQNFVVAARSQSDNFYVSEKTDTPMIGTLAGANSVLNCPVEVGHTEHLNKSQARQIKNNPKFLVLPQESQKRAVA"

112..984

/product="put. mature Ala subunit (aa 1-291)"

985..1539

/product="put. mature Bx subunit (aa 1-185)"

1714..1719

/note="put. polyadenylation signal"

1746

/note="polyadenylation site"

BASE COUNT

554 a

418 c

399 g

375 t

ORIGIN

Query Match

Best Local Similarity

27.4%;

Score 508;

DB 8;

Length 1746;

62.0%;

Pred No. 4.4e-112;

Db 24 ATGGCTTAAGCTTCTTGACATTTCTTCATCTGTTTCTGCTTTTCAG---TAGCTGC 80
Qy 61 atctccttcagcagcagcgcggaggaatcgtgcagttccagcgcctcaatgcgcag 120
Db 81 TTTGCTCTAAGAGAACAGTCTCAACAAATAGAGTGCAGCTCGAAGCATCAATGCTCTC 140
Qy 121 agactgacaacgcgcatgaatcgagggcggttacatgagactgggaaccccaacac 180
Db 141 GAGCTGATACCGTATAGAGTGGGAAGTGGTCTCATTTGAGACATGGAATCCCAACAC 200
Qy 181 caggagtgcgaatgc 240
Db 201 AGGCAATTTGATGCGCGGTGGCGCTCTCTCGTGTACCTTTCAACGAAATGCCCTT 260
Qy 241 cgtaggccttctactcaatgcctccagagagatcttcacagcaagcaagggagatac 300
Db 261 CGGAGACCTTATTTCCATATGCTCCCAAGAAATTTATCCCAACAAAGTATGGATAT 320
Qy 301 ttgggttgatattccctggttctctagcacatatgaagagcctgcacacaagaagcgc 360
Db 321 TTTGGCATGTTTCCCTGTTGCTGTGAGACCCATGAGGAGCCACAACAA----- 371
Qy 361 cgatcatgcccaagaacaccacagcgttgcaagaagaagacacaaagcacaagcaa 420
Db 372 -----TCTGAACAGGAGAGGAGCAGCGGTAC 398
Qy 421 caagatgaccaggaaggtgcaccgtttcaatgaggggtgatctcattcagttccacc 480
Db 399 AGAGAGACCCACCAAAAGGTTAACCGATTACAGAGAGGGGTGATATTCAGTTCCTACT 458
Qy 481 ggtgttcttctggtgtacaaagcagacacagcagatgatttctgtttcttcttaact 540
Db 459 GGTATTGCAATTTGGATTGTAACAGCAAGATACCTCCAGTTATTGCCAATTTCTCTTACC 518
Qy 541 gacacacaaacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 600
Db 519 GACACTGGCAGCTCCATAACACAGCTTGTATGATGCTAGGAGATTCTATCTGTGGG 578
Qy 601 aaccagcagcagaggttcttaaggtaccagcagcagcagcagcagcagcagcagcagc 560
Db 579 AACCAAGAGCAAGAGTCTACGATACAGCAGATCAA-----CAAGGAGGAAGCAAGA 631
Qy 661 ttaccataagccatacagccgcagcagcagcagcagcagcagcagcagcagcagcagc 720
Db 632 CCAAGACATGATGGGACACATTTTCAGTGGCTTCAAGAGGATTTCTTGAAGATGC 691
Qy 721 cctcagagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 780
Db 692 TTTCAACGTGAACAGGCATATAGTAGACAGACTTCAAGGCGAGGAAGACGAGAGAA 751
Qy 781 aacatcttcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 840
Db 752 GGGAGCCATTTGCAAGTGAAGGTGCTCAGCATCATAGCTCC----- 796
Qy 841 cagattgtcaaaatctgtgggcgagcagcagcagcagcagcagcagcagcagcagcagc 900
Db 797 -----ACCGAGAGGCAAGCAGCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 843
Qy 901 gtgagggagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 960
Db 844 ACGAAGATGAAGAGGAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 893
Qy 961 gaatcagatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatga 1020
Db 894 -----GATGAGAGCAG 947
Qy 1021 agagggcgggggaatggttattgaagagacgatctgcacccagcagcagcagcagcagcatt 1080
Db 948 AGGCAAGAGAGAGCAATGGGCTTTGAAGAAACCGTTTGCACTGCCAACTTCCGGGCAACAT 1007
Qy 1081 ggtggaacacagatcccttcacatcagatcctcagcagcagcagcagcagcagcagcagc 1140

Db 1008 GGTTCATCTCCGTACACAGACATCTACAACCTCAAGCTGGTAGAATCAAAACTGTTACC 1067
Qy 1141 gatcacaaccttaatacttaagtggtgagcttagtgagctgaatggaatggaatggaatggaat 1200
Db 1068 AGCCTTGACCTCCAGTTCAGGTGGCTCAAACTAAGTGGCTGAACATGGATCTCTCCAC 1127
Qy 1201 aggaatgcatgtttgtccctcactacaacacacacacacacacacacacacacacacacacac 1260
Db 1128 AAAACGCTATGTTGCGCTCACTACAACCTCAACGCAACAGTGTATATATACGCATTG 1187
Qy 1261 agggagcggcctcagctgcaagtggtagacgcaacgcaacgcaacgcaacgcaacgcaacgcaac 1320
Db 1188 AAGGAGCGTCAAGGGTACAAATGTTCAACTGCAATGGCAACACTGTGTTGATGGAGAG 1247
Qy 1321 cttaacaggggtcagcttctgtggtgcccacagaaacttcgcggtggtggtggtggtggtggtggt 1380
Db 1248 CTAGAAGCGGTGCTGATGATGAGTGCAGTGCACAAACATATGCAGTGGCTGCTAAGTCATTG 1307
Qy 1381 agcagaaacttcgaatacgtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggt 1440
Db 1308 AGCAGAGGTTCAATATGTTGGCATTTCAAGACGGATGATAGAGCTAGTATTCGAAGACTT 1367
Qy 1441 gccggtgaaactccttcctcatagataacctgcggagggaggtggtggtggtggtggtggtggtggt 1500
Db 1368 GCAGGACATCATCAGTTATAGATGATCTGCTTTGATGATGATGATGATGATGATGATGATGATGAT 1427
Qy 1501 ctccagggagc 1560
Db 1428 ATGAGAGGAATAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1487
Qy 1561 cctttcagcagctcctccagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1594
Db 1488 CCAGCCAGTCTGAGATGAGAGCTTCGGCTTAGA 1521

RESULT 15

VFLEA2
LOCUS VFLEA2 1666 bp mRNA PLN 18-MAY-1995
DEFINITION Vicia faba mRNA for Legumin A2 pre-pro-polypeptide.
ACCESSION X55014
VERSION X55014.1 GI:22007
KEYWORDS legumin A; storage protein.
SOURCE Vicia faba
ORGANISM Vicia faba
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Viciaeae; Vicia.
REFERENCE 1 (bases 1 to 1666)
AUTHORS Bassuener, R.
TITLE Direct Submission
JOURNAL
REFERENCE 2 (bases 1 to 1666)
AUTHORS Schlesier, B., Bassuener, R., Nong, V. H. and Muentz, K.
TITLE The cDNA derived primary structure of two distinct legumin A subunit precursors from field bean (Vicia faba L.)
JOURNAL Nucleic Acids Res. 18, 7146-7146 (1990)
MEDLINE 91088307
REFERENCE 3 (bases 1 to 1666)
AUTHORS Horstmann, C.
TITLE New aspects of the subunit structure of Vicia faba legumin
JOURNAL Kulturpflanzen 32, 109-116 (1984)
REMARK (sites)
REFERENCE 4 (bases 1 to 1666)
AUTHORS Bassuener, R., Wobus, U. and Rapoport, T. A.
TITLE Signal recognition particle triggers the translocation of storage globulin polypeptides from field beans (Vicia faba L.) across mammalian endoplasmic reticulum membrane
JOURNAL FEBS Letters. 166 (2), 314-320 (1984)
MEDLINE 84108944
PUBMED 6198216

BASE COUNT	537 a	359 c	385 g	385 t
ORIGIN	/genes="Lea2"			
Query Match	23.7%; Score 439.8; DB 8; Length 1666;			
Best Local Similarity	59.3%; Pred. No. 1.3e-95;			
Matches 946; Conservative	0; Mismatches 557; Indels 92; Gaps			

24 ATGCTAAACTCTTGCACTCTCTCTTTCACATTGCTTCTACTTTTCAG ---TAGCTGC 80

Db	24	ATGGCTAACTTCTTGCACTCTCTCTTTTCCATCTTGCTTTCTACTTTTCAG---TAGCTGC	80
QY	61	atctccttcaggcagcagcgaggagagaatgcgtgcaggttccagccctcaatgcgcag	120
Db	81	TTTGCTCTCAGAGAACAGTCTCAACAAAATGAGTCTAGCTAGAACGCCTCGATGCCTTT	140

[illegible]

Qy 241 cgtaggcctttctactccaatgctccccaggagatctttcatccagcaaggaggatcac 300

[illegible]

Qy 361 cgatacagtcccaaagaccacccaagagctttgcagaagaagaccacccaagcccaacgcaa *420
||| ||| | | | | |
Db 372 -----TGTGAACAGAGAGAGGGAGCGGCTAC 398

421	QY	caagatagtaccagaaggtgcacccgtttcaatgagggtgatctcaatgcagttcccacc	480
399	Db	AGAGACAGCCACCAAAAGGTTAACCGATTAGAGAGGGTGATCATTTGCAGTTCCTACT	458
481	QY	gggtgtgcttcttgctgctgaacagaccacgacactgattgttgcgtttctcttact	540
459	Db	GGTATTGTATTTGGATGTCAAATGACCAAGATATCTCCAGTATTGCCATTTCCTTACT	518
541	QY	gacaccaacaacagcacaccagcttgatcagttcccccaggagattcaatttggctggg	600

[illegible]

Qy 661 ttaccatatagccatacagcccgcatagtcggccctagacgagaagcgtgaatttcgc 720

Db 635 AGACAATGACGGGAACAACATTTTCAGCGGCTTCAAGAGGATTTCTTTGGAGATCGGTT 694

	721	cctcggaggacagccacagccgcgacaggaacgacgaggaacaaga - gaagaagacgaagggggg	/ /
Qy			
	695	GAACGTGAACAGGCATATAGTAGACAGACTTCAAGGCCGAGAACCGAAGACGAAGAAGGG	754
Db			

QY 780 aaacatcttcagggtcacgcggagttcctggaacaagccttcagggttgacgacag 839

Db 755 AGCCATTGTCAAAG--TGAAGGTGGTCTCAGCATCATCAACCCCTCCGGAGAGGCAA--- 809

840	QY	acagattgtgc	aaatctgtg	ggcgagac	gagagtg	gaagaag	ggagccatt	ggac	
810	Db	-----	-GCGGCCAC	CCCGAGG	CGCAGAC	AGCAAGA	AAAGAT	GAAGAT	CAAAA
900	QY	ggtgagg	gagcctc	agaattct	tgacc	cagatg	gaacag	agaggt	ccgcacaa
863	Db	-----	-AGACAAC	CAAGCC	ATCAATA	AAGCAG	CAAGAG	GTGAAG	ATCAAGA

Db 863 GGAAGAG-----AGACAAACCAAGCCATCATAAAGCAGAAGAGGTTGAAGATGAAGA 913

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 27, 2001, 13:04:00 ; Search time 8470.79 Seconds
(without alignments)
3957.396 Million cell updates/sec

Title: US-09-715-036-5

Perfect score: 2032

Sequence: 1 aataatcatatattatc.....cgttgtgcgtgttttctcc 2032

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb.ba.*

2: gb.htg.*

3: gb.in.*

4: gb.cm.*

5: gb.ov.*

6: gb.pat.*

7: gb.ph.*

8: gb.pl.*

9: gb.pr.*

10: gb.ro.*

11: gb.sts.*

12: gb.sy.*

13: gb.un.*

14: gb.vi.*

15: em.ba.*

16: em.fun.*

17: em.hum.*

18: em.in.*

19: em.om.*

20: em.or.*

21: em.ov.*

22: em.pat.*

23: em.ph.*

24: em.pl.*

25: em.ro.*

26: em.sts.*

27: em.sy.*

28: em.un.*

29: em.vi.*

30: em.htgo_hum.*

31: em.htgo_inv.*

32: em.htgo_rod.*

33: em.htg_hum.*

34: em.htg_inv.*

35: em.htg_rod.*

36: em.htg_other.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

8

Result No. Query Match Length DB ID Description

1	2032	100.0	2032	6	AX148742	AX148742 Sequence
2	2032	100.0	2032	6	ARQARAH	L34402 Arachis hyp
3	2019.2	99.4	2032	6	AX155331	AX155331 Sequence
4	1805.2	88.8	1978	6	AX155332	AX155332 Sequence
5	1752.4	86.2	1949	8	ARQARAH	L38853 Arachis hyp
6	465.8	22.9	1433	8	PSVIC	Y00722 Pisum sativ
7	461	22.7	1561	8	PSVIC	Y00462 Vicia faba
8	449	22.1	1596	8	PSVIC	X67429 Pisum sativ
9	435.4	21.4	1548	8	VNVICLN	Z71987 V.narbonens
10	408.4	20.1	1251	8	ABO08679	ABO08679 Glycine m
11	404	19.9	1632	8	ABO08678	ABO08678 Glycine m
12	404	19.9	1955	8	GMASBC	X17698 Soybean mRN
13	402	19.8	1680	8	ABO08680	ABO08680 Glycine m
14	378.8	18.6	2069	8	PSA276875	AJ276875 Pisum sat
15	347.8	17.1	1858	8	VNCONVN	Z71988 V.narbonens
16	339.4	16.7	1494	8	GMU14006	U14006 Glycine max
17	336.2	16.5	1086	8	PEAVIC7	J01260 pea vicilin
18	304.6	15.0	1534	8	CECANAV	X05733 Canavalia g
19	303.2	14.9	907	8	PEAVIC2	J01258 pea vicilin
20	303	14.9	1403	8	CECANAV	X59467 C.ensiformi
21	250.4	12.3	1035	8	SOY7SAA	J01290 soybean 7s
22	239.6	11.8	1047	8	SOY7SNPA	J01294 soybean 7s
23	204.4	10.1	794	8	SOYBCONGL	M26128 Soybean bet
24	178	8.8	530	8	SOY7SAB	J01291 soybean 7s
25	178	8.8	936	8	SOY7SAC	J01292 soybean 7s
26	175.4	8.6	1404	8	U01122	U01122 Phaseolus l
27	174.2	8.6	1535	8	PVPHASAR	X02980 Phaseolus v
28	172.2	8.5	1478	8	U01121	U01121 Phaseolus l
29	171.4	8.4	1741	8	LCVICNE	Z48440 L.culinaris
30	167	8.2	1341	8	AF348366	AF348366 Cajanus c
31	160.6	7.9	591	8	PSCONVR1	X01379 Pea mRNA fr
32	160.2	7.9	2140	6	AX031079	AX031079 Sequence
33	160.2	7.9	2141	8	AF161885	AF161885 Macadamia
34	160	7.9	1742	8	LCVICNA	Z48436 L.culinaris
35	160	7.9	1743	8	LEVICN	Z48438 L.ervoides
36	157.4	7.7	2171	6	AX031077	AX031077 Sequence
37	157.4	7.7	2182	8	AF161884	AF161884 Macadamia
38	157	7.7	4239	8	PSVICIL	X14076 Pea vicilin
39	155.6	7.7	1475	6	A11822	A11822 Artificial
40	154	7.6	1454	8	U01131	U01131 Phaseolus v
41	154	7.6	1475	8	PVPHASBR	X03004 Phaseolus v
42	154	7.6	5327	8	VFVICG	Y00506 Vicia faba
43	152.8	7.5	1430	8	U01132	U01132 Phaseolus v
44	152.6	7.5	2171	6	AX031075	AX031075 Sequence
45	152.6	7.5	2182	8	AF161883	AF161883 Macadamia

ALIGNMENTS

RESULT 1
AX148742
LOCUS AX148742 2032 bp DNA
DEFINITION Sequence 5 from Patent WO0136621.
ACCESSION AX148742
VERSION AX148742.1 GI:14347296
KEYWORDS
SOURCE peanut.
ORGANISM Arachis hypogaea

PAT 08-JUN-2001

REFERENCE
AUTHORS Dodo,H.W., Arnzen,C.J., Konan,K.N. and Viquez,O.M.
TITLE Down-regulation and silencing of allergen genes in transgenic peanut seeds
JOURNAL Patent: WO 0136621-A 5 25-MAY-2001;
FEATURES Alabama A & M University (US)
Location/Qualifiers

Arachis hypogaea

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Aschynomeneae; Arachis.

1 (bases 1 to 2032)


```
Qy 961 gagcagccgacacatcctactctgagggcttcagcagagaatacgttgagagcc 1020
Db 961 GAGCAGCCGACACATCTACTTTCAGGAGATTCAGCAGGAATACGTTGAGGCCGC 1020
Qy 1021 cttcaatgcgaattcaatgagatcacgaggggtgctgttagaagagaatgcaggggtga 1080
Db 1021 CTTCAATGCGAATTCAATGAGATACGAGGGTGTGTTAGAAGAGAATGCAGGAGGA 1080
Qy 1081 gcaagagagagagagcagagggcgatgagctcggagtgtgagaaacaataaagact 1140
Db 1081 GCAAGAGAGAGAGAGGAGGAGGATGGAGTACTCGAGTAGTGAGAAACAATGAAGAGT 1140
Qy 1141 gatagtcgaagtgtcaagaggagcagcttggaagaacttactaaacacgctaaatccgtctc 1200
Db 1141 GATAGTCAAGTGTCAAGGAGCAGCTTCAAGAACTTACTTAAGCAGCTTAAATCCGTCTC 1200
Qy 1201 aaagaagagctccgaagaagagggagatatcaccaccccaatacaacttgagagaagcga 1260
Db 1201 AAAGAAGAGCTCCGAAGAAGAGGGAGATATCACCAACCCCAATCAACTTGAGAGAAGCGA 1260
Qy 1261 gccqgactttctaaacactttgggaagttatttgaggtgaagccagacaagaagaaccc 1320
Db 1261 GCCGATCTTCTACAACTTGGAGAGTTATTGAGGTGAAGCCAGACAGAAAGAACCC 1320
Qy 1321 ccagcttcagagacctgacatgatgtcactctgttagagatcaagaagagagctttgat 1380
Db 1321 CCAGCTTCAGGACCTGGACATGATGCTCACTGTGTAGAGATCAAAAGAGGAGCTTTGAT 1380
Qy 1381 gctccacacttcaactcaagggccatggttatctcgctgctcaacaagaagaaactgaaa 1440
Db 1381 GCTCCACACTTCAACTCAAGGGCCATGGTTATCGTCTGCTCAACANAGAACTGGAAA 1440
Qy 1441 ccttgaactctgtgtgaagaagagcacaacagagagggagcgcggggaagaagaga 1500
Db 1441 CTTTGAACCTGTGTGAAGAAGAGCAACAACAGAGGGGAGCGCGGGAAGAAGAGA 1500
Qy 1501 ggaagagacacaagaagaggaagtaacagagaggtgctgagtacacagcagaggt 1560
Db 1501 GGACGAACAGCAAGAGAGAGGAGGAATACAGAGAGGTGCTGTAGGTACACAGCGAGGT 1560
Qy 1561 gaaggaagcgatgtgttcattcatctgcagcagctcatccagtagccatcaacgcttcctc 1620
Db 1561 GAAAGGAAGCGATGTGTTCATCATGCGCAGAGCTCATCCAGTAGTCAACGCTTCTCTC 1620
Qy 1621 caactccatctgttgcttcggtatcaacgctgaacacagacagagatcttccttc 1680
Db 1621 CAACTCCATCTGTGCTTCGGTATCAAGCTTGAAACCAACACAGAACTCTCCTTGC 1680
Qy 1681 aggtgataagacaatgtgataccagatagagaagcgaagcgaagattagcattccc 1740
Db 1681 AGGTGATAAGGACAATGTGATAGACCAGATAGAGAAGCAAGCGAAGGATTTAGCATCCC 1740
Qy 1741 tgggtcgggtgaacaagttgagaagctcatcaaaaacagaagaatctcaatttgag 1800
Db 1741 TGGGTGCGGTGAACAAGTTGAGAAGCTCATCAAAAACAGAAGAACTCTCACTTTGTGAG 1800
Qy 1801 tgctgctcaaatctcaatctcctgctgtctcctgagaaagagtcctcctgagaa 1860
Db 1801 TGCTGCTCAATCTCAATCTCAATCTCGTCTCTCTCTGAGAAAGAGTCCTCTGAGAA 1860
Qy 1861 agaggatacaagagaggaacacaaagagaggaaggggtccactctttcaatttgaaggc 1920
Db 1861 AGAGGATCAAGAGGAGGAAAAACCAAGGAGGGAAGGGTCCACTCTCTTCAATTTGAAGGC 1920
Qy 1921 ttttaactgagaatgagcgaactgttatgtatcgtataaagatacaagcctttgtact 1980
Db 1921 TTTTAAGTGAATGAGGCAACTTGTATGTATTCGATATAAGNATCACGCTTTTGTACT 1980
Qy 1981 ctactatcaaaaacttataataataaataaagctttgtcgtttgtctcc 2032
Db 1981 CTACTATCAAAAACCTTATCAATAATAAATAAAGCCTTTGTGCTGTTTCTCC 2032
```

```
RESULT 4
AX155332 1978 bp DNA PAT 22-JUN-2001
LOCUS Sequence 5 from Patent WO0140264.
AX155332
VERSION AX155332.1 GI:14536767
KEYWORDS
SOURCE
peanut.
ORGANISM
Arachis hypogaea
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae;
Aeschynomeneae; Arachis.
REFERENCE
1 (bases 1 to 1978)
Bannon,G.A., Burks,W.A., Caplan,M.J., Sampson,H. and Sosin,H.
Peptide antigens
AUTHORS
TITLE
JOURNAL
FEATURES
Location/Qualifiers
source
1..1978
/organism="Arachis hypogaea"
BASE COUNT 614 a 460 c 513 g 391 t
ORIGIN
Query Match 88.8%; Score 1805.2; DB 6; Length 1978;
Best Local Similarity 95.2%; Pred. No. 0;
Matches 1935; Conservative 0; Mismatches 43; Indels 54; Gaps 5;
Qy 1 aataatcatatatattcatcaatcatctataagtagtagcaggagcaatgaggggag 60
Db 1 AATAATCATATATATTTCATCAATCATCTATATAAGTAGTAGCAGGAGCAATGAGAGGGAG 60
Qy 61 ggtttccactgagctgtgtctgaggtaccttgcctggcttcagttctgcaacgca 120
Db 61 GGTTCCTCCACTGATGTGTGTCTGAGGGATCCTTGTCTGGCTTCAGTTCTGCAACGCA 120
Qy 121 tgcaagtcataccttaccagaagaacacagagaacccctgcgccagaggtgcctcca 180
Db 121 GGCCAAGTCAACC-----TTACCGGAAACAGAGAACCCCTGGCGCCAGAGGTGCCTCCA 174
Qy 181 gagttgtcaacagaaacccgatacacttgaagcaaaaggcatgcgctcgctgacacaa 240
Db 175 GAGTTGTCAACAGGAACCCGGATGACTTGAAGCAAAAGGCAATGCGAGTCTCGCTGCACCAA 234
Qy 241 gctcagtgatgactcctgctgtgtctatgatcctcagaggacacactggcaccacca 300
Db 235 GCTCGAGTATGATCCTGTTGTGTCTATG-----ACACTGGCGGCCACCAACCA 282
Qy 301 acgttccctccagggagggagacacgtgcccacaaacccggagactacgataacccg 360
Db 283 ACGTCACCTCCAGGGAGCGGACACGTGGCGCCCAACCCGGAGACTACGATGATGACCG 342
Qy 361 ccgtcaaccccgaaagagagagggcgcgatgggacacagctggaccagggagcgctga 420
Db 343 CCGTCAACCCCGAAGAGAGAGGAGGAGGCCCGCATGGGGACCAAGCTGAACCGAGGGAGCGTGA 402
Qy 421 aagagaagaagactggagagacaacaaagagaagattggaggcgaccagaagtcacagagcc 480
Db 403 AAGAGAAGAAGACTGGAGAGAACCAACCAAGAAAGATTGGAGCGGACCAGTCAAGTCAAGCAGCC 462
Qy 481 acggaataaaggcccgaaagagagaggaaggaacaaagagtggggaacaccaggtagcca 540
Db 463 ACGGAATAAAGGCCCGAAGGAAGAGAGAGGAACAAGAGTGGGGAAACACCAAGGTAGCGGA 522
Qy 541 tgtagggagaagaacatctcgggaacacccctttctactccctcgaagcggttttagcac 600
Db 523 GGTGAGGGAAGAAACATCACGGAACAACCCCTTTCTACTTCCCGTCAAGGCGGTTTAGCAC 582
```

QY 601 ccgctacgggaacaaacggttagatccggctcctgcagaggtttgacaaaggtcaag 660
 Db 583 CCCTACGGGAACCAAAACGGTAGGATCCGCTCTCTGAGAGGTTTGACCAAAAGGTCAA 642
 QY 661 gcagtttcagaaactccagaaacacccgtattgtgcagatcgaggccaaacccctaacactct 720
 Db 643 GCAGTTTCAGAACTCCAGAAATCCAGATCCAGGTATTGTCAGATCGAGGCCACACCTAACACTCT 702
 QY 721 tgtcttcccaagcagcgtgtagtgcataacatctctgttatccagcaagggcagccac 780
 Db 703 TGTCTCTCCCAAGCAGCGCTGATGCTGATTAACATCTCTGTTATCCAGCAAGGACAGCCAC 762
 QY 781 cgtgacgttagcaaatgccaataacagaaagagcttcaatcttgacgagggccatgcact 840
 Db 763 CTGTACCGTAGCAAAATGCGCAATACAGAAAGAGCTTTAATCTTGACGAGGGCCATGCACT 822
 QY 841 cagaatcccatccggtttcattcttcatactcttgaccgcgcacgacacccagaacctcag 900
 Db 823 CAGAAATCCCATCCGTTTCATTTCTTACATCTTGAATCGCAATGCAACACCAAGAACCTCAG 882
 QY 901 agtagctaaatctccatcccggttaacacacccgcccagtttgagatttctcccgcc 960
 Db 883 AGTAGCTAAATCTCCATGCCCGTTTAAACACGCCCGCCAGTTTGAGGATTTCTTCCCGCC 942
 QY 961 gacgagccgagaccaaatacctactctgagggcttcagcaggaatacgtttggaggccg 1020
 Db 943 GAGCAGCCGAGACCAATCATCTACTTGCAGGATTCAGCAGGAATACTTTGGAGCGCG 1002
 QY 1021 ctccaatgcgaattcaatgagatacagaggtgctgttagaagaagatgcagaggtga 1080
 Db 1003 CTTCAATGCGGAATCAATGAGATACGAGGCTGCTGTAGAAAGAGATGACGAGGAGA 1062
 QY 1081 gcaagaggagagggcgacagggcgtgaggtactcgtgagtagtgagaaacatgaagaggt 1140
 Db 1063 TCAAGAGGAGAGGCGCAGCGCAGGAGTACTCGGAGTAGT-----TAAAGAGAGT 1117
 QY 1141 gatagcaagtgtcaaaaggacagcgttgaaagaactactaagcgcgttaactcgtctc 1200
 Db 1118 GATAGTCAAAAGTCAAAAGGACGACGCTTCAAGAACTTACTAAGCAGCGTAATCCGTCTC 1177
 QY 1201 aaagaaggtccgaaagagagggagagatcaccaaccaatcaacttgagagagggcga 1260
 Db 1178 AAGAAAGGCTCGAAGAGA---AGATATCAACCAACCACTCACTTGAGAGATGGCGA 1234
 QY 1261 gccgatcttttaacaaacttgggaagtatttgagtgaggccagacaagaagaaccc 1320
 Db 1235 GCCCGATCTTCTAACAACCTTTGGAGGTATTGAGGTGAAGCCAGACAAAGAAGAACCC 1294
 QY 1321 ccaggtccaggaacctgagatgatgctcaactgtgtagagatcaagaagagagttgat 1380
 Db 1295 CCAGCTTCAGGACCTGGATGATGCTCACTACCTGCTGTAGAGATCAAGAAGGAGCTTTGAT 1354
 QY 1381 gctccacacacttaactcaaaagccatggtattcgtcgtcgtaacaaaggaaactggaaa 1440
 Db 1355 GCTCCACACTTCAACTCAAAAGCCATGCTCATCTGCTGCTCAACAAGGAACCTGGAAA 1414
 QY 1441 ccttgaaactgtggtgtgaagaaagagacaacacagaggggacggcgaggagagagga 1500
 Db 1415 CTTTGAACCTGTAGCTGTGAAGAAAAGAGCAACAACAGAGGGGACGCGGGGAACAGAGTC 1474
 QY 1501 ggcgaagacgaagaagagaggggaagttaacagagaggtgcgtaggtgtcacagcgaggtt 1560
 Db 1475 GGACGAAGCAGAGAGAGAGGGAAGTAACAGAGAGGTGCGTAGTACACAGCGAGGT 1534
 QY 1561 gaaggagggcgaagtgttcatcatgccagcagctcatccagtagccatcaacgttcttc 1620
 Db 1535 GAAGGAAGCGAGTGTGTTTCATGCTGCAAGCAGCTCATCCAGTAGGCATCAACGCTTCTC 1594
 QY 1621 cgaactccatctgctgtggttcgttatcaacgctgaacaaacacacacgaatcttcttcg 1680
 Db 1595 CGAACTCCATCTGCTTGGCTTCGGTATCAACGCTGAAACCAACACCAAGATCTTCTTGC 1654
 QY 1681 aggtgataaggacaaatgtgatagaccagatagagaagcaagcaagaggttagcattccc 1740

Db 1655 AGGTGATAGACCAATCTGATACAGATAGAGAAAGCAAGCGAGGATTTAGCATCCCC 1714
 QY 1741 tgggtcgggtgaaacaggttgagaagctcatcaaaacacagaaagaaatctcactttgtgag 1800
 Db 1715 TGGTTCCGGTGAAACAGTTGAGAAGCTCATCAAAACAGAGGAGTCTCACTTTGTGAG 1774
 QY 1801 tgcgtcctcctaactcctaactcctaactcctcgtcgtcctcctgagaaagaggtctcctgagaa 1860
 Db 1775 TGCTCCTCTCAATCTCCGT-----GTCTCTCTGAAAA 1806
 QY 1861 agaggtcatcaagagaggaacaaacaaagggaggggtccactcccttcaattttgaaggg 1920
 Db 1807 AGAGGATCAAGAGGAGGAAACCAAGAGGAGGAGGTCCTCTTCAATTTGAAGGC 1866
 QY 1921 ttttaactgagaatggaggcaactgttatgtatgcataataagatcacgcttttgtact 1980
 Db 1867 TTTTAACTGAGAAATGGAGGAAACTTGTATGTATCATCAATAAAGATCACGCTTTGTAAAT 1926
 QY 1981 ctactatcaaaaacttatcaataataaaaacgtttgtggtgtttctcc 2032
 Db 1927 CTACTATCCAAAACTTATCAATAATAAAACGTTTGTGCGTTGTTCTCC 1978

RESULT 5
 AROARAH 1949 bp mRNA PLN 10-JAN-1995
 LOCUS Arachis hypogaea (clone P17) Ara h I mRNA, complete cds.
 DEFINITION L38853
 ACCESSION L38853
 VERSION L38853.1 GI:620024
 KEYWORDS peanut hypersensitivity.
 SOURCE Arachis hypogaea (strain Florunner) Seed cDNA to mRNA.
 ORGANISM Arachis hypogaea
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae;
 Aeschynomeneae; Arachis.
 REFERENCE 1 (bases 1 to 1949)
 AUTHORS Burks, A.W., Cockrell, G., Stanley, J.S., Helm, R.M. and Bannon, G.A.
 TITLE Recombinant peanut allergen Ara h I expression and IgE binding in patients with peanut hypersensitivity
 JOURNAL Unpublished
 FEATURES
 Location/Qualifiers
 1..1949
 /organism="Arachis hypogaea"
 /strain="Florunner"
 /db_xref="taxon:3818"
 /dev_stage="Seed"
 1..2
 /base="Ara h I"
 /gene="Ara h I"
 3..1847
 /gene="Ara h I"
 /codon_start=1
 /protein_id="AAA60336.1"
 /db_xref="GI:620025"
 /translation="MRGRVSPMLLLILVLASVSATQAKSPYKRTENPCAQRCLOSC
 QEPDLDKAKCESRCTKLEYDPRCTVDTGATNQHPGRTGRQPGDIDDDRRQPR
 REEGRWGPAEPREEREEDWRQDPREDWRPSHQPKIRPEGREGEWGTGSEVR
 EETSRNPFYFSRRFSTRYQNQRIRVLQRFQDRSKQFQNLQNHRIQVIEARPNL
 VLPHADANLIVTQQGQATVTVANGNNRKSFNLDGHALRIPSGTISYILNRDNN
 LRVAKISMPVNTPGQDFPFPASSRDQSSYLQGSFRTNLEAFNAEPNERRRLLBN
 AGGQEPERGQRRTSRSDNEGVTKVSEHVEQLTKHAKSKVSKSEEDINPIL
 KNGEPDLNNEGRFLFEVKPKKNPQLQDLDMMLTCVEIKEGALMLPHNSKAMVIVV
 RDGTNLELVAVRKQEQQRGEWEDEEEDDEEENREVRRTARLKEGVDVFTMP
 AAPHVAINASSELHLLGFGINAENNRHIFLAGDKDNVDIQIEKQADLPAPGSGEOVE
 KLINQRESHFVSARPOSQSPSPSEKEDQEENQGGKGPLLLILKAFN"
 1848..1949
 3'UTR
 polyA_site 599 a 455 c 517 g 378 t
 BASE COUNT
 ORIGIN

Query Match 86.2%; Score 1752.4; DB 8; Length 1949;
Best Local Similarity 95.0%; Pred. No. 0;
Matches 1894; Conservative 0; Mismatches 46; Indels 54.

QY	48	caatgagaggaggggtttctccactgatgtgtgtctagggataccttgcctggcttcag	107
Db	1	CAATGAGAGGAGGGGTTCTCCACTGATGCTGTGCTGGATCTCTTGCTTGGCTTCAG	60
QY	108	ttcttgaacgcattgcgaagtcatcaacttaccagaagaacacagaaacccctgcgcc	167
Db	61	TTTTCTGTGCAACGACGGCCAGTAC-----TTACCGGAAAACAGAAACCCCTCGGCC	114
QY	168	agagtgctccagagtttcaacaggaacccggtgacttgaagcaaaagcatcgagt	227
Db	115	AGAGTGCTCCAGATTGTCAACAGGAACCGACGACTGAAGCAAAAGGCATCGGAGT	174
QY	228	ctcgctcaccaagctcgagtatgatactcgtgtgtgtctatgatcctcgtgagacactg	287
Db	175	CTCGCTGACCAAGCTCGAGTATGATCTCTGTGTGCTCTATG-----ACACTG	222
QY	288	gcacacaaaccaagttccctccaggggagcggacgtgcccgcgaaccccgagact	347
Db	223	CGCCACCAACCAACGTCACCTCCAGGGAGCGGACACGTGCCGCCAACCCGGAGACT	282
QY	348	acgtatgacccgcgttcaaccccgagagagaagggccgattgggaccagctggac	407
Db	283	ACGATGATGACCGCGGTCAACCCCGAAGAGAGGAAGGCGGATGGGACCAGCTGAAC	342
QY	408	cgaggagcgtgaaagagaagaagactggacacacaaagagaagattggaggcgacca	467
Db	343	CGAGGGAGCTGAAAGAGAAGAAGACTGGAGACACCAAGAGAAGATTGGAGGCGACCA	402
QY	468	gtcatcgacgcaccggaataaagcccggaaggagaagagagaaacagagtggggaa	527
Db	403	GTCAATCAGCACCACGGAATAAAGCCCGAAGGAAGAGAAGAGAACAGAGTGGGAA	462
QY	528	caccagatgacatgtagggaagaacatctcgaaacaccccttctacttcccgtaaa	587
Db	463	CACCAAGTACGAGGTGAGGGAAGAACAATCAGGAACAACCTTTCTACTTCCCGTCAA	522
QY	588	ggcggttttagcacccgtctacgggaaccaaaccggtagatccgggtcctcagaggttg	647
Db	523	GGCGGTTTAGCACCCGCTACGGGACCAAAACGCTAGGATCGCGCTCTGCAGAGGTTTG	582
QY	648	acaaaggccaaggagtttcagaaatctccagaaatcaocggtattgtgcagatcagagcca	707
Db	583	ACCAAGGTCANAAGCAGTTTCAGAAATCTCCAGAAATCCGCTATTGTGCAGATCGAGGCA	642
QY	708	aacttaacctgttcttcccaagcacgtgctgataaactcttgtattatccagc	767
Db	643	GACCTAACACTCTGTCTTCCCAAGCAGCTGATGCTGATAACAATCTGTTATTCAGC	702
QY	768	aaggccaagccacggttacgtagcaaatggcaataacagaaagagctttaatcttgacg	827
Db	703	AAGGACAAGCCACCGTGACGTAGCAAAATGGCAATAACAGAAAGAGCTTTAATCTTGACG	762
QY	828	agggcattcaactcagaatcccatccggtttcaattctcatatttgaacgcataaaca	887
Db	763	AGGCCCATGCACTCAGAATCCCATCCGTTTCATTTCCTACATCTTGAATGCAGATGACA	822
QY	888	accagaacctcagatgagtaaaatctccatcccggttaacacaccccgccagtttgagg	947
Db	823	ACCAGAACCTCAGAGTAGCTAAAATCTCCATGCCGTTAAACACGCCCGCCAGTTTGAGG	882
QY	948	atttcttccggcgagcagccagaccaatactacttctgaggggttccagcagggaata	1007
Db	883	ATTTCCTCCGGGACGACCCGAGACCAATATCTACTTTCAGGGGATTCAGCAGGAATA	942
QY	1008	cgttgaggccgccttcaatcgsgaatccaatgataacgagaggtctgttagaagaga	1067
Db	943	CTTTGGAGCCGCTTCATGCGGGAATCAATGAGATACGGAGGGTGCTGTTTAGAAGAGA	1002
QY	1068	atgcaggaggttagcaagaggagagaggcagagcgatgagttactcgaqtaqtaaa	1127

VERSION	Y00722.1	GI:20915	
KEYWORDS	storage protein; vicilin.		
SOURCE	pea.		
ORGANISM	<i>Pisum sativum</i>		
ORGANISM	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Viciaeae; <i>Pisum</i> .		
REFERENCE	1 (bases 1 to 1433)		
AUTHORS	Watson, M.D.		
TITLE	Direct Submission		
JOURNAL	Submitted (25-JAN-1988) Watson M.D., Dept. Botany, Science Laboratories, Univ. of Durhathi, South Road, Durhathi DH1 3LE		
REFERENCE	2 (bases 1 to 1433)		
AUTHORS	Watson, M.D., Lambert, N., Delauney, A., Yarwood, J.N., Croy, R.R., Gatehouse, J.A., Wright, D.J., and Boulter, D.		
TITLE	Isolation and expression of a pea vicilin cDNA in the yeast <i>Saccharomyces cerevisiae</i>		
JOURNAL	Biochem. J. 251 (3), 857-864 (1988)		
MEDLINE	88362626		
FEATURES	Location/Qualifiers		
source	1..1433		
	/organism="Pisum sativum"		
	/strain="Feltham First"		
	/db_xref="taxon:3888"		
	/clone="pDUB9."		
sig_peptide	<1..8		
CDS	<1..1304		
	/codon_start=3		
	/product="vicilin precursor"		
	/protein_id="CAA68708.1"		
	/db_xref="GI:758248"		
	/db_xref="SWISS-PROT:P13918"		
	/translation="SSRSDDQNPFFKSNKFTQFLFNENGHIRLLKQFDQRKIFNLQNYLLEYKSPHFIPLQHTDADYILVLSGKAILTVLXPDNRNFNLERGDTIKLPAGTAYLVNRDNEELRVLDLAI PVNRPGLQSLSGNOQNYLSGFSKNILKASFNTDVEIEKVLLEHEKETOHRSLKDKROSOBNVIVKLSRGQIEELSKNAKTSKKSVSSEPEPNLRSRGPITYSNEGKPEITPEKNPOLDLIDFVNSVEIKESGLLPHYNSRAIVTVNEGKDFELVGNORNEQORKEDEEPEQEEENKQVQNYKAKLSGDVFPVAPGHPVAKASSNLDLLGFGINAENRNPLAGDENIVISQVRPKELAFPGSAQEVDIRLENNQKSHFDAQPPQQRERSRETRDLSSV"		
mat_peptide	9..1301		
	/product="vicilin"		
BASE COUNT	508 a 259 c 283 g 383 t		
ORIGIN			
Query Match	22.9%;	Score 465.8;	DB 8; Length 1433;
Best Local Similarity	62.5%;	Pred. No. 3e-118;	
Matches 785;	Conservative 0;	Mismatches 452;	Indels 18; Gaps 3;
QY 559	tcggaacaacccctttctactctcccgcaagcggttttagcaccgcgtacggaacccaaaa 618		
Db 17	TCTCAAAATCCTTTTACTCAAGCTTAACAAGTTCAAACTCTTTTGAGAAATGAAA 76		
QY 619	cggtaggataccgggttcctgcagaggtttgaccacaaaggtcaaggcaggttcagaatccca 678		
Db 77	TGGGCACATTCGACTCTCGCAAAATTTGACCAAGTTCTAAAATTTTCGAGAAATCTACA 136		
QY 679	gaatcacggtattgagatcgagcgacaaacctaaacctttgttctcccaagcagc 738		
Db 137	AACTACCGCTCTTTTGGAAATATAAGTCCAAACCTCACAAATATTTCTCCACAGCAC 196		
QY 739	tgatgctgaatacactcttcttatccagcaagggcagccaccgtgacctagcaaatgg 798		
Db 197	CGATGCCGATTTACATCTTGTGTACTCAGTGGAAAGCTATATCATCAGGTGTGAAACC 256		
QY 799	caataacaagaagacttaattcttgcagagggccactgcactcagaatcccatcggttt 858		
Db 257	CGATGATAGAACTCTCTCAACTTTGACGCGGAGATACGATAAATCTCTGCTGCGAC 316		
QY 859	catttctacacttgaacccgcacatgacaacacagaaacctcagatagctaaatctccat 918		
Db 317	AAATTGCTTATTGGTTAAACAGAGATGACAAAGAGAGCTTAGACTATTAGATCTCGCCAT 376		
QY 919	gccgttaacacaccccgccagtttgaggattttctcccgcgagcagcgacacaaac 978		
Db 377	TCCCGTAAATAGACCTGGCAACTTCAGTCTTCTTATTTGTGAAATCAAAACCAACA 436		
QY 979	atcctacttgcagggtcttcagcaggaatacgttgaggccgccttcactcgggaattcaa 1038		
Db 437	AAACTACTTATCTGGGTTTCAGTAAGAACAATTTCTAGAGGCTTCCTTCAATCTACTATGA 496		
QY 1039	tgagatacggaggtgctgttagaagagaatgcagggagtgagcaagagagagagggca 1098		
Db 497	AGAGATAGAAAAGGTTCTTTTAGAGAGCATGACAAAGACACACACAGAGAGAGCCT 556		
QY 1099	ggggcgatggaggtactcggagtagtgagaacaatgaaggagtagtagtcaaaagtccaaa 1158		
Db 557	TAAGGATAAGAGGCGCAAGAACTCAAGAAGAGAAAT-----GTAATAGTAAATATCAAG 610		
QY 1159	ggagcacgttgaagaacttactaagcacgcgttaaatccgtctcaagaagaaggtcccgaga 1218		
Db 611	GGGACAAATTTAGGAAATTTGAGTAAAAATGCAAAAGTCTACCTCCAAAAAAGTGTTCCTC 670		
QY 1219	agaggagatatacaccaccccaatcaacttgagagaagggcgagcccgatcttttctaaca 1278		
Db 671	TGAATCTG-----AACCATTCNACTTGAGAAGTCGCGGTCTCTATCTATTCCACGA 721		
QY 1279	ctttgggaagtatttgggtgaagccagacaagaagaacccccagcttcagagacctgga 1338		
Db 722	GTTTGGAAATTTCTTTGAAATCACTCCAG---AGAAAAATCCACAGCTTCAAGACTTGA 778		
QY 1339	catgatgctcacctgttagagatcaagaagagctttgatgctcccccacactcaactc 1398		
Db 779	TATATTTGCAATTCGTAGAGATTAAGGAGGATCTTTATTGTTGCCACACTACAAATTC 838		
QY 1399	aaagccatggttctcgtcgtcacaacaaaggaaactggaaacctggaactcgtggtctg 1458		
Db 839	AAGGCCATAGTAATAGTAACAGTTAACGAAGGAAAGAGAGATTTTGAACCTTGGGTCA 898		
QY 1459	aagaaaagacacaacagagagggcgggggaagaagagagagagagagagagagagaga 1518		
Db 899	AAGAATGAAACCAACCAAGAGACAGAGAAAAGAGATGACGAGGAGAGAGAAAGAGAGA 958		
QY 1519	ggagggaagtacagagaggtgctgttaggtacacagcgaggttgagggaagcgatggtt 1578		
Db 959	AGAAGAGATAAATAACAAAGTGCAAAATTTACAAAGCTAAATTTATCTTTCAGGAGATGTTT 1018		
QY 1579	catatgcagcagctcatcagtagccatcaacgcttctccggaacctccatctggttgg 1638		
Db 1019	TGTGATTCACAGAGCCCATCCAGTTGCCCTAAAAGCTTCCTCAAAATCTTGATTGCTTGG 1078		
QY 1639	cttcggtatcaacgctgaaacacacagaaatcttcttgcaggtgataagagacaatgt 1698		
Db 1079	GTTTGGTATTAACTGTGAGACATCAGAGGAACTTTCTTGACGGCGATGAGCATATGT 1138		
QY 1699	gatagaccagatagagaagcaagcgaaggatttagcatccctcgggttcgggtgaaacaagt 1758		
Db 1139	GATTAGTCAGGTACAGCACCGAGTGAAGAGCTTGCATTCCTCGATCAGCTCAAGAGGT 1198		
QY 1759	tgagagctcatcaaaaaacacagaggaatctcactttgtgagtgctcgtctccaa 1813		
Db 1199	TGATAGGATATAGAGATCAGAAACAAATCCACATTTGCAGATGCTCAACCTCAA 1253		
RESULT 7			
VFVIC	1561 bp	mRNA	PLN
LOCUS	Vicia faba mRNA for vicilin storage protein.		11-MAY-1995
DEFINITION	Vicia faba mRNA for vicilin storage protein.		
ACCESSION	Y00462		
VERSION	Y00462.1	GI:22052	
KEYWORDS	7S seed globulin; storage protein; vicilin.		
SOURCE	fava bean.		
ORGANISM	Vicia faba		
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Viciae;
Vicia.

REFERENCE 1 (bases 1 to 1561)

AUTHORS
TITLE
JOURNAL

Direct Submision

Submitted (11-NOV-1987) Bassuener R., Akademie der Wissenschaften
der DDR, Zentralinstitut fuer Genetik und Kulturpflanzenforschung,
Corrensstrasse 3, DDR - 4325 Gatersleben, Germ. Dem. Rep

REFERENCE 2 (bases 1 to 1561)

AUTHORS
TITLE

Bassuener, R., Hai, N.V., Jung, R., Saalbach, G. and Muntz, K.
The primary structure of the predominating vicilin storage protein
subunit from field bean seeds (Vicia faba L. var. minor cv. Fribo)
Nucleic Acids Res. 15 (22), 9609 (1987)

88067789

see also Y00506 [Nucl. Acids Res. 15:10065-10065(1987)] for

corresponding genomic sequence.

Data kindly reviewed (12-FEB-1988) by Bassuener R.

FEATURES

Location/Qualifiers

1..1561

/organism="Vicia faba"

/variety="minor"

/cultivar="Fribo"

/db_xref="taxon:3906"

/clone="pvi 91"

23..103

23..1414

/codon_start=1

/product="vicilin precursor"

/protein_id="CAA69525.1"

/db_xref="GI:22053"

/db_xref="SWISS-PROT:P08438"

/translation="MAATLKDSFLLTLILGIAFLASVCLSSRDNDNPFVFNRFQ
TLFENENGIRLQKFDQHSKLELNLQVRLLEYKSKPTFLPQTDADFLVLSG
KAILVLPNDRNSFLSFGDFIKLPACTIGLYNRDDEBLRDLVLPVNRPEGP
SFLLSGNQNPISLGSFNKLEASFNIDYKEIKVLLHEHGEKHYHRRGLDRRQG
QENVIVLSKQIELELNKNAKSSKSTSESEFNLKRSPIYSNKFGRFFETTPK
RNPQLQDNLVFNVEINEGSLLPHNSRAIVIVVNEGKDFELVGNRNEOGLR
EYDEKQGEIEIRKQVNYKAKLSPGDLVLPAGYPVAKASSNLNLVGFNAEN
NORYFLAGEDNVIQIHKPKVELAFPGSAQEVDFLLENQKSHFANNAQPOEREGSG
EIKDLYSLIGSF"

mat_peptide

104..1411

/product="unnamed"

1470..1480

/note="putative"

1498..1508

/note="putative"

1527..1540

/note="putative"

BASE COUNT 547 a 287 c 298 g 429 t

ORIGIN

Query Match

Best Local Similarity 22.7%; Score 461; DB 8; Length 1561;

Matches 782; Conservative 0; Mismatches 455; Indels 18; Gaps 3;

QY 559 tcggaacaacccctttactctccgccaaggcggttagcaccgcctcaggaacaaaa 618

DB 115 TCAAGATAACCTTTTGTCTTTGAGTCTAACAGGTTTCAAACTCTTTTGAGATGA 174

QY 619 cggtagatccgggtctcagaggtttgaccacaagggtcaaggcagtttcagaatccca 678

DB 175 CGGTACATTCGTCTCTCCAAAAATTTGATCAACATTCCAAATCTTGTGAGATCTCCA 234

QY 679 gaatcaccgtattgagatcgagggcaaacctaacactctgttcttcccaagcagc 738

DB 235 AAATFACCGTCTTTTGGATATAAATTAACCTTCAACCATATTTCTTCCACAAAC 294

QY 739 tgatgctgataacatccttctgttatccagcaaggcgcaagcccggtgacgtagcaaatgg 798

DB 295 TGACGCCGATTCATCCTTGTGTCTTAGTGGAAGAGCTATCTACACAGGTATTATACC 354

QY 799 caataacagaagaagagctttaattcttgacgagggccatgcactcagaatcccatccggttt 858

DB 355 CAACGACAGAAACTCCTTCAGTCTTGAGCGAGGAGATACCAATAAACTTCCTGCTGGTAC 414

QY 859 catttccatcatcttgacccgcatgacaaccagaacacacacacacacacacacacacac 918

DB 415 AATTTGGTTATTTAGTTAAACCGAGATGACGAGGAGGATCTTAGAGTACTAGATCTTTGTCA 474

QY 919 gcccttaacacacaccccgccagtttgaggatttctcccgccgagcagccgagacacacac 978

DB 475 CCCCGTAATAGACAGGTGAACCTCAGTCTTTCTTTATTTGTTGTTGTAATCAAAACCAACC 534

QY 979 atctacttgcaggcttcagcaggaatacgttggagccgcttcaatcggaatccaa 1038

DB 535 GTCCATCTTATCTGATTACGAACAACATTTTAGAGGCTTCTTCAATACCGATTATAA 594

QY 1039 tgatatacggagggtgctgttagaagaatacagaggttgagcaagagagagagagagagca 1098

DB 595 AGAGATAGAGAGGTTCTCTTAGAAGAGCATGGGAAAGAGAAATATACACAGAGAGGCTT 654

QY 1099 gaggcgatggagctactcgagtagtgagaacaatgaaggagtagtagtcaagtgtcaaa 1158

DB 655 TAAAGATAGGAGCGAGCGGGCCAAAGAGAAAT-----GTAATAGTCAAAATATCAAG 708

QY 1159 ggagcacgttaagaacttactaagcacgctaaatccgctcctcaagaagaagcgtccgaaga 1218

DB 709 GAACAAATGAGGAATTTGAACAAAAATGCAAAAGTCTAGCTCCAAAAAGAGATACATCTTC 768

QY 1219 agaggagatatacacaacccaatcaacttgagagaagcgagccgctattcttaaca 1278

DB 769 TGAATCTGAA-----CGGTTCAATTTGAAAGTCTGAGGCCCATCTATTCCAACAA 819

QY 1279 ctttgggaagttatttgaggtagagccagacaagaagaccccgcttcaggacctgga 1338

DB 820 STTTTGGTAAATTTTGTAGATCACCCCA---AAGAAATCCCAACTTCAAGACTTGAA 876

QY 1339 catgatgctcacctgtgtagatcaagaaggagctttagtgcctccacactcaactc 1398

DB 877 TATATTTGCTCAATATGTGGAATTAATGAGGATCTCTTTTGTGGCACACTACAATTC 936

QY 1399 aaaggccatggtatcgctcgctcaacaagggaactggaaaccttgaaacctcggtgtg 1458

DB 937 AAGGCCATAGTAGTAAACAGTCAATGAAGAAAGGGGATTTTGAACCTTGTGGGTCA 996

QY 1459 aagaaagagcaacaacagaggggacggcggggaagagagagagagagagagagagagag 1518

DB 997 AAGAAATGAGAACCAACAGGCGCTGAGAGAAATATGACGAGGAAAGGAAGAGAGA 1056

QY 1519 ggagggaagtacaagagaggtgctgtaggtacacagcagaggttgaggaagcagatggtt 1578

DB 1057 AGAAGATAGGAACAAAGTGCAAAATTCAAAGCTTAAGTTGTCCTCAGGTGATGTTT 1116

QY 1579 catatgccagcagctcatccagtagccatcaacgcttctcccgaaactccatctgcttg 1638

DB 1117 AGTAATTCAGCGGGTTATCCAGTTGCCATAAAAGCTTCATCAATCTCAATTTGGTGG 1176

QY 1639 cttcggtatcaacgctgaacaaacacacacacacacacacacacacacacacacacac 1698

DB 1177 ATTTGGCATCAATCTGAGAACACACAGAGATATCTCTTTCAGGTGAGGAAGACAATGT 1236

QY 1699 gatagacagatagagaagcaagcgaaggtattagcattccctgggtcggtgaacaagt 1758

DB 1237 GATAAGTCAGATACATTAACACGTAAGAAACTTGCATTCCTTGGATCAGCTCAAGAGT 1296

QY 1759 tgagaagctcatcaaaaacacaggaaggaatctcactttgtgagtgctgctgctcaaa 1813

DB 1297 AGATACGCTACTAGAGAAATCAGAAACAATCTCACTTTGCAAAATGCTCAACCTCAA 1351

RESULT 8

PSVICK

LOCUS

DEFINITION Psium sativum vick gene.

PSVICK 1596 bp mRNA

PLN 29-OCT-1997

ACCESSION X67429
 VERSION X67429.1
 KEYWORDS GI:297169
 SOURCE 47 kDa protein; seed storage protein; vicillin; vick gene.
 ORGANISM Pisum sativum
 Eukaryote: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Viciae;
 Pisum.
 REFERENCE 1 (bases 1 to 1596)
 Bown, D.P.
 Direct Submission
 TITLE Submitted (02-JUN-1992) D.P. Bown, University of Durham, Dept. of
 JOURNAL Biological Sciences, Science Site, South Rd, Durham, DH1 3LE, UK
 REFERENCE 2 (bases 1 to 1596)
 Bown, D.P.
 Thesis (1992) Durham University
 REFERENCE 3 (bases 1 to 1596)
 Thompson, A.J., Bown, D.P., Yalish, S. and Gatehouse, J.A.
 Differential expression of seed storage protein genes in the Pea...
 JOURNAL Biochem. Physiol. 187, 1-12 (1991)
 FEATURES
 Location/Qualifiers
 1..1596
 /organism="Pisum sativum"
 /cultivar="Feltham first"
 /db_xref="taxon:3888"
 /clone="pUG1.63"
 /tissue_type="cotyledon"
 /clone_lib="plasmid pUC8"
 /dev_stage="13-17 days post flowering"
 13..1329
 /gene="vick"
 13..1329
 /gene="vick"
 /codon_start=1
 /product="vicillin 47kD protein"
 /protein_id="CAA47814.1"
 /db_xref="GI:297170"
 /db_xref="SPTREMBL:O43626"
 /translation="MAATPKLMLLAFLASVCVSSRDQENPFKSNRFTLYE
 NENGHILKQNFDRKSRKIFENLQNYRLLEYKSPHILFLQYTDADFLVLSKATL
 TVLKSNDNRNFERGDAIKLPAGTIAIYLRNDDNEDLRVIDLAIAPVNFQGLQSL
 SGTONQPSLLSFKNTLEAFNTNIEIEKVLLEQOEQPHRRSLKDRRQINBEEN
 VIVKVSREQIEELSKNAKSSKSSVSEFPNLRNPYISNKKFKFEITPEKNOQ
 LODLIDFVNSDIKESGILLPNYNSRAIVITVTGKDFELVQNRNENQKNDKEE
 EOETSKVOLYKRAKLSPGDVFIIPAGHPVAINASSDLNLIGFVNAENNERFLAG
 EEDNVISOVERPKVELAPGSSHEVDRLNQRKQSYFANAQPLQRE"
 1552
 polyA_site 586 a 302 c 283 g 425 t
 BASE COUNT
 ORIGIN
 Query Match 22.1%; Score 449; DB 8; Length 1596;
 Best Local Similarity 62.5%; Pred. No. 1.4e-113;
 Matches 783; Conservative 0; Mismatches 440; Indels 30; Gaps 4;
 QY 559 tcggaacaccctttctactccctccagcgggttttagcaccctcagcgggaacaaaa 618
 Db 93 TCAGAGAACCCCTTTATCTTTAAGTCTAACCGATTCAAACTCTTATGAGAACGAAAA 152
 QY 619 cggtaggataccgggttcctcagaggtttgacaaagggtcagggttcagaaatctcca 678
 Db 153 CGGTCACTACGCTCTCTCCAAAAATTTGACAAAGGTTTCCAAAAATTTGAAAAATCTTCA 212
 QY 679 gaataccattgtgcagatcgaggcaaacctaacactctgttctcccaagcagc 738
 Db 213 AATTACCGCTCTTTAGATATATAGTCCAAACCTCACACCCCTTTCTCCCAATACAC 272
 QY 739 tgatgctgaatacactctgtttatccagcaaggcgaacccgtgacagtagcaaatgg 798
 Db 273 CGATGCCGACTTCATCTGTGTAGTCTTGTGGGAAAGCCACTCTCACAGTGTGGAATC 332
 QY 799 caataacagaagagctttaattgacgagggccatgcactcagaatccctcggttt 858

Db 333 TAACGATCGAAACTCCTTCAATCTTGAACGTTGGTATGCCATCAAACTCCCTGCTGGCAC 392
 QY 859 catttccatactctgaaccgccatgaacaacagacactcagatgtagtataaactccat 918
 Db 393 TATGCTTATTTGGGTAAACCCGAGATGACAAAGAGATCTTAGAGTATTAGATCTGCCAT 452
 QY 919 gccgttaacacacccggcaggtttgaggtttctcccgccgagcagccgagacaaac 978
 Db 453 CCAGTAAACAACCCGGTCAATTGCAGTCTTCTTATTATATGGAACCTCAAAATCAACC 512
 QY 979 atctactctgagggtctcagcaggaaatcgttgaggccgcttcaatgcgggaattcaa 1038
 Db 513 ATCATTATTATCTGGATTGACGAAGAAATTTCTAGAGCTGCTTTCAATACCAATFACGA 572
 QY 1039 tgagatacggagggtgctgttagaagagaatcagagggtgagcaaggagagagggcga 1098
 Db 573 GGAGATAGAAAAGGTTCTTTTAGAACACAGGAACACAGCCACACAGAGAAAGTCT 632
 QY 1099 gaggcgatggagtactcggagttagtgagaacaatgaaggagtgtagtagtaagtgtcaa 1158
 Db 633 TAAGGATAGGACAAAGAGA-----TCAACGAAGAAAAATGTAATAGTCAAAAGTATCAAG 686
 QY 1159 gagcagctgtgaagaacttactaagcacgctaaatccgtctcaagaagaaggctccgaaga 1218
 Db 687 GGAACAATTTGAGGAATTGAGCAAAATGCAAGTCCAGTTCACAAAAAAGTGTATCATC 746
 QY 1219 agaggagatatcaccacacccaactcaactgagagaaggcggcgcgtatcttcttaacaa 1278
 Db 747 AGAATCTGG-----ACCATTCACTTGAAGTCCGCAATCTATCTATCTAACAA 797
 QY 1279 ctttgggaagttattgaggtgaagccagacagaagaagaccccgcttcaggaacctgga 1338
 Db 798 GTTTGGCAAAATCTTTGAGATCACCCAG---AGAAAAATCAACAACCTTCAAGACTTGA 854
 QY 1339 catgagctcacctgtgtagagatcaagaagagctttgagctccacactcaactc 1398
 Db 855 CATATTGTCAATCTGTGGATTATGAGGGGATCTTTTATTGTTGCCAAACTACAAATTC 914
 QY 1399 aaaggccatggttatcgtcgtcgctgcacaaagaagcgtggaaccttgaactcgtggtgt 1458
 Db 915 AAGAGCAATTTGTATAGTAACCTTTACCGAAGGAAAAAGGAGATTTTGAACCTTGTGGTCA 974
 QY 1459 aagaagaagcaacacagagggggcggcggggaagagagagagagagagagagagagaga 1518
 Db 975 AAGAAATGAGAACCA-----GGGAAAAAATAATGACAAAGAAAGAGGAACAA 1022
 QY 1519 gggggaagttaacagagaggtgctgtaggtacacagcagaggttgaaggaggcagatgtt 1578
 Db 1023 ARAAGACAAACAAACAGTGAAGCTATATAGAGCTAAGTTGTTCTCCAGGTGATGTTT 1082
 QY 1579 catcatgccagcagctcattccagtagccatcaacgcttctccgaactccatcgtttgg 1638
 Db 1083 TGTGATTCAGCAGGTGACCCGCTTGCCATAAATGCTCCTCAGATCTCAATCTGATTGG 1142
 QY 1639 cttcgggtatacagcgtgaac 1698
 Db 1143 ATTTGGTATCAATGCCGAGAACACGAGAGAAATCTTCTTGCAGGTGAGGAACACATGT 1202
 QY 1699 gatagacagatagagagcaagcagaggatttagcattccctcgggtcgggtgaaacaagt 1758
 Db 1203 CATAGTCAAGTAGAAGACCAGTTAAAGAGCTTGAAGAGTTCCTTCCCTGGATCTTCTCATGAGGT 1262
 QY 1759 tgagaagctcatcaaaaaacacagagaaggaatctcactcttctgtgagtcgctcctc 1811
 Db 1263 TGATAGGCTCCTAAGAAATCAAAACAACTCTTTTGTGCAAAATGCTCAGGCCTC 1315

RESULT 9
 VNWICLN VNWICLN 1548 bp mRNA PLN 30-APR-1996
 LOCUS V.narbonensis mRNA for vicilin.
 DEFINITION
 ACCESSION Z71987

LOCUS	AB008679	1251 bp	mRNA	PLN	01-SEP-2000
DEFINITION	Glycine max mRNA for beta subunit of beta conglycinin, partial cds.				
ACCESSION	AB008679				
VERSION	AB008679.2	GI:9967358			
KEYWORDS	beta subunit of beta conglycinin.				
SOURCE	Glycine max (cultivar:wasesuzunari) cDNA to mRNA.				
ORGANISM	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.				
REFERENCE	1 (sites)				
AUTHORS	Maruyama,N., Katsube,T., Wada,Y., Oh,M.H., Barba De La Rosa,A.P., Okuda,E., Nakagawa,S. and Utsumi,S.				
TITLE	The roles of the N-linked glycans and extension regions of soybean beta-conglycinin in folding, assembly and structural features				
JOURNAL	Eur. J. Biochem. 258 (2), 854-862 (1998)				
MEDLINE	99089652				
REFERENCE	2 (bases 1 to 1251)				
AUTHORS	Maruyama,N., Katsube,T., Wada,Y., De La Rosa,A. and Utsumi,S.				
TITLE	Direct Submission				
JOURNAL	Submitted (04-NOV-1997) Shigeru Utsumi, Kyoto University, Research Institute for Food Science, Gokancho, Uji, Kyoto 611, Japan (E-mail:utsumi@soya.food.kyoto-u.ac.jp, Tel:0774-38-3760, Fax:0774-38-3761)				
COMMENT	On Sep 5, 2000 this sequence version replaced gi:2605511.				
FEATURES	Location/Qualifiers				
source	1..1251				
	/organism="Glycine max"				
	/cultivar="wasesuzunari"				
	/db_xref="taxon:3847"				
	<1..1251				
	/codon_start=1				
	/product="beta subunit of beta conglycinin"				
	/protein_id="BAA23361.2"				
	/db_xref="GI:9967359"				
CDS	<p> NIRYRIVQSPKPTLLPHADADFLFLVLSGRAILLVNDDRDVSNLHPDQAE IPAGTYYILNPHDHLNLIKALPVNKPGRVDDPFLSTQAQSYVLOGFSNILET SFSEFEINLPLFGEERQRQGVIVELSKQIROLRSRKSRRKTISSEDEPN LRSNPILYNNFGFFETPEKNPQLRDLDFLSSVDINEGALLPHFNKALIVLI NEGDANIELYIKQKQOEPELVEVRYRAELSEDDVFIPIAAYPFVYNATSLN FLATGINAENNRNFLAGEDNVVYQIERQVQELAFPGSAQVERLLKKQREYFVDA PQPKQSGSKGRKGPSPILGLV" </p>				
BASE COUNT	385 a	304 c	265 g	297 t	
ORIGIN					
Query Match	20.1%; Score 408.4; DB 8; Length 1251;				
Best Local Similarity	62.2%; Pred. No. 2.7e-102;				
Matches	780; Conservative	0; Mismatches	381; Indels	93; Gaps	5;
QY	562	gaacaaacctttctactt---cccgtcaaggcggttttagcacccgctacgaggaacaaaa	618		
Db	21	GAATACCCCTTTCTACTTTAGAGCTTAACAGCTTCCAAACTCTCTTTGAGAACCAAAA	80		
QY	619	cggtaggataccgggtcctgcagaggtttgacaaagggtcaaggcaggtttcagaatctcca	678		
Db	81	CGGTGCGATTGCTCTCTCCAGAGATTCAACAAACGCTCCCCACAACTTGAGAACCTTCG	140		
QY	679	gaataccgattatgtgcagatcgaggccaaaccttaacactctgttcttcccaagcagcgc	738		
Db	141	AGACTACCGGATTGTCAGTTTCAGTCAAAACCCCAACAACTCTCTCCCCACCATGC	200		
QY	739	tgatgtgataaacatcttggttatccagcaagggaagccagccgctgagcagtagcaaatgg	798		
Db	201	TGACGCCGATTTCTCTCTTGTCTTAGCGGAGAGCCATCTTACCTTGGTGAACAA	260		
QY	799	caataacagaagaagcctttaaattcttgacagaggccatcgaatcccatcccggttt	858		
Db	261	CGACGACAGAGACTCTTACAACCTTACCCCTGGCGATGCCAGAGAAATCCAGCTGGAAC	320		
QY	859	catttctcatcttgaaaccgcatgacaaacacgaacctcagtagtctaaatctccat	918		
Db	321	CACTTACTATTGTTTAAACCCCTCAGCAGCACAGAAATCTCAAAATAATCAAACTTGCAT	380		
QY	919	gccggttaacacacccggccagtttgagagatttctccggcgagcagccgagacaatc	978		
Db	381	ACCCGTCAAACAAACCTGCGAGATATGATGATTCTTCTTATCTAGCATCAAGCCCAACA	440		
QY	979	atcctacttgagggtcttcagcaggaatacgttgaggccgcttcaatgcggaattcaa	1038		
Db	441	GTCTTACTTTCGAGGCTTCAGCCATAATATTCTAGAGACCTCTCTTCCATAGCAATTGCA	500		
QY	1039	tgagatcacggagggtgtctgtttagaagagaatcgagaggtgagcaagaggagagggca	1098		
Db	501	GGAGATAAACAACAGGTTTCTTTGGAGAGAGAGAGGAGAGGAGAGAGAGAGAGAGAG	552		
QY	1099	gaggcgatgagtagtactcggagtagtgagaaacatgaagagtagtagtcaaaagtcaaa	1158		
Db	553	-----GGAGTGATGCTCGGAACACTTCAAA	575		
QY	1159	ggagcagcttgagaagaacttactaagcacgctaaatccgtctcaaaagaaggctccgaaga	1218		
Db	576	GGACAAATTCGCAACTCAGCAGAGCTGCCAAATCTAGTTCAAGGAAAACCATTTCTCTC	635		
QY	1219	agaggagatatacccaaccataacttgagagaaaggcgagcccgatcttttatacaaa	1278		
Db	636	CGAAGATG-----AACCATTCAACTTGAGAGCGCAACCCCATCTATTTCACAAACA	686		
QY	1279	ctttgggaagtatttgaggtgaagccacagaagaagaacccccagcttcagagacctgga	1338		
Db	687	CTTTGGAAAGTTCTTTGAGATCACCCCTG---AGAAAAACCCACAGCTTCGGGACTTGA	743		
QY	1339	catgatgctcacctgtgtagatcaaaaggagcgtttgatgctcccacattcaactc	1398		
Db	744	TACTTCTCCAGTCTGTGGATATCAACGAAGGAGCTCTTCTTACCACACTTCAATTC	803		
QY	1399	aaaggccatggttatcgtcgtcgtcaacaaaggaaactggaaccttgaactcgtgctgt	1458		
Db	804	AAAGGCCATAGTGATCTACTAGTGATTATGAAGGAGATGCAACATTTGAACCTTGTGGCAT	863		
QY	1459	aagaaaaagacaacacagcgggacgagcggggaagagagagagagagagagagagaga	1518		
Db	864	TAAAGAACACACACAGACGAAACAGAGAGAACCTTTG-----	906		
QY	1519	ggagggaagtaacagagaggtgctgtaggtacacagcagaggttgaaagagagcgatgctt	1578		
Db	907	-----GAAGTGCAAAAGGTACAGAGCTGAATTTGCTCTGAAGACGATGATT	950		
QY	1579	cataatgccagcgctcatccagtagccatcaacgcttctccgaactccatctgctgtg	1638		
Db	951	TGTAATTTCCAGCAGCTTATCCATTGTCTCAACGCTACCTCAAACTCAATTTCTCTGC	1010		
QY	1639	cttcggtatacacgctgaaacacacacagaatcttcttcgaggtgataagagacaatgt	1698		
Db	1011	TTTTGGTATCAATGCTGAGAACACACAGAGAACTTCTTTCGAGCGGAGAGAACATGT	1070		
QY	1699	gatagacagatagagagcaagcaagggatttagcattccctgggtcggtggaacaagt	1758		
Db	1071	GGTAAGGAGATAGAAAGACAAAGTGCAGGAGCTTGCCTTCCCTGGGTCTGCACAAGATGT	1130		
QY	1759	tgagaagctcatcaaaacacagaggaatctcactttgtgagtgcctgcctca	1812		
Db	1131	TGAGAGGCTATTAAAGAGCAGAGGGAATCTTACTTTGTTGATGCTCAGCCTCA	1184		
RESULT	11				
LOCUS	AB008678	1632 bp	mRNA	PLN	01-SEP-2000
DEFINITION	Glycine max mRNA for alpha subunit of beta conglycinin, partial cds.				
ACCESSION	AB008678				
VERSION	AB008678.2	GI:9967356			
KEYWORDS	alpha subunit of beta conglycinin.				
SOURCE	Glycine max (cultivar:wasesuzunari) cDNA to mRNA.				

alpha' subunit of beta-conglycinin.
 Glycine max (cultivar:Wasesuzunari) cDNA to mRNA.
 Glycine max
 ORGANISM
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
 Glycine.
 REFERENCE
 1 (sites)
 Maruyama, N., Katsube, T., Wada, Y., Oh, M. H., Barba De La Rosa, A. P.,
 Okuda, E., Nakagawa, S. and Utsumi, S.
 The roles of the N-linked glycans and extension regions of soybean
 beta-conglycinin in folding, assembly and structural features
 Eur. J. Biochem. 258 (2), 854-862 (1998)
 99089652
 2 (bases 1 to 1680)
 Maruyama, N., Katsube, T., Wada, Y. and Utsumi, S.
 Direct Submission
 Submitted (04-NOV-1997) Shigeru Utsumi, Kyoto University, Research
 Institute for Food Science, Gokancho, Uji, Kyoto 611, Japan
 (E-mail:utsumi@soya.food.kyoto-u.ac.jp, Tel:0774-38-3760,
 Fax:0774-38-3761)
 On Sep 5, 2000 this sequence version replaced gi:4191813.
 COMMENT
 Location/Qualifiers
 source
 1..1680
 /organism="Glycine max"
 /cultivar="Wasesuzunari"
 /db_xref="taxon:3847"
 <1..1680
 /codon_start=1
 /product="alpha' subunit of beta-conglycinin"
 /protein_id="BA74452.2"
 /db_xref="GI:9967361"
 /translation="VEEEEEEGQIPRPHQPHRERQHQKEEDGEGPRPFPFP
 RQPHQHEHQKEHEHWRKEHKGSEEDERHPRPHOPHOKEEHKEHWH
 KQKHQKSEEEEDQDEQKESQESQEPFRHKNKPNFNSKRFQTL
 FKQYGHVRVLFNKRQQLNDRYLFENSKPNTLLPHADLIVILNGTA
 ILVLVNDSDVSNQDALRVAGTVYVVPNDENLRMITLAIPVKNPGRFES
 FLSTQAOQSLQGSFKNTLEASDTPKEEIKVLFREGOQOGEERLQPSVIEIS
 KQIRELSKHSKSKSTLSSDQPNLRSRDPYISNKLKFLFTPKNQLRDLDV
 FLVDMNEGALFLPHFNKAIWLVINEGEANIELVGIKEQQOQQOGEERLEVRKY
 RAELSDDIFVPIAGTVVFNATSDNFPAFGINAEINRNFLAGSKDNVLSIQPSQV
 QELAFPGSKADINELIKSESIFYDVAQPOQKEGNKGRKGPLSLRAFY"
 BASE COUNT
 ORIGIN
 567 a 385 c 395 g 333 t
 Query Match 19.8%; Score 402; DB 8; Length 1680;
 Best Local Similarity 61.8%; Pred. No. 1.7e-100;
 Matches 770; Conservative 0; Mismatches 405; Indels 75; Gaps 5;
 Qy 563 acaaacctttctactccgcgaagcggttttagcacccgcgtacggaaacaaacggt 622
 Db 439 AAGAACCTTTTCACTCAACTTAAAGGTTCCAACTCTCTTCAAAAACCAATATGCG 498
 Qy 623 aggtatccgggttcctgcagaggttgacaaaggtcaaggcagtttcagatatccagaat 582
 Db 499 CACGTTCCGCTCCCGAGAGGTTCACAAACCGCTCCCAACAGCTTCAGAACTTCCGAGAC 558
 Qy 683 caccgatgtgcagatcgaggccaaacaccttaacactctttgttcccaagcacgctgat 742
 Db 559 TACCGCATTTGGAGTTCACCTCCAAACCCACACACCTTCTTCCCCACCACATGCTGAC 618
 Qy 743 gctgataacacttcctgttatccagcaaggaggccaccgctgaccctgacccatagcaaatgcaat 802
 Db 619 GCTGATTACCTCATCGTTATCTTAAAGGAGCTGCCATTTCTTACCTTGGTGAACAACAGAC 678
 Qy 803 aacagaagaagacttaattcttcagcagggccatgcactcagaatccatccggtttcatt 862
 Db 679 GACCGAGACTCTTACACCTTTCAATCTGGCGATGCCCTTAAGAGTCCCTCGAGGAACACA 738
 Qy 863 tctacacttgaaccgcctatgacaacacagaaacactcagatagctaaatctctgcc 922
 Db 739 TACTATGTGGTTAACCTTGACACGACGAGAGTAATCTCAGATGATGATACACTCGGCATACCC 798

Qy 923 gtttaacacacccgccagtttaaggatttcttcctcggcgagcagccgagacccaatcatcc 982
 Db 799 GTTAACAAACCCCGGTAGATTGAGAGTTTCTTCTATCTAGCAGCTCAAGCTCAACAGTCC 858
 Qy 983 tacttgagggtcttcagcaggaatcagttggaggcgccctcaatgcggaattcaatgag 1042
 Db 859 TACTTGAAGGTTTCAGCAAGAATAATCTAGAGCGCTCATAGACACCAAAATTCAGGAG 918
 Qy 1043 atacgagggtgtgttagaagaagaatcagaggtgagcaagaagagagagagggcagagg 1102
 Db 919 ATAAACAGGTTCTGTTGTAGAGAGGAGGGCACCAACAGGGGAGAGGCTG--- 975
 Qy 1103 cgaatggagtactcgagtagtgagacaatgaaggagtagtgatcaaatgtcaagagg 1162
 Db 976 -----CAAGAGAGTGTGATTGTGGAATCTCAAGAAA 1008
 Qy 1163 cagttgaagaacttactaagcagcgtataatccgtctcaaaagaaggctccgaaagaag 1222
 Db 1009 CAAATCGGGAACGTGAGCAAAACATGCAATCTAGTTCAAGGAAAACCAATTTCTTCTGAA 1068
 Qy 1223 ggagatatcaccaccccaatcaacttgagagagcgagcccgatcttcttaacaacttt 1282
 Db 1069 G-----ATAACCTTTCAACTTGAGAGCGCGACCCCATCTATTCCACAAGCTT 1119
 Qy 1283 ggaagttatttgaggtgaagccagacagaagaacccccagcttcagacactgacatg 1342
 Db 1120 GCGAAGTGTGTTGAGATTACCCAG---AGAAAAACCCCTCAGCTTCGGGACTTGGATGTC 1176
 Qy 1343 atgctacactgtgtagagatcaagaagagcgtttgatgctcccaactcaactcaag 1402
 Db 1177 TTCTCAGTGTGTTGATATGACAGGGGAGGCTCTTTTCTGCCACACTCAATTCAAAG 1236
 Qy 1403 gccatggtatcgctcgctcgcatacaagaagaaactggaacactggaactcggtgtaaga 1462
 Db 1237 GCCATAGTGTACTAGTGTATTAATGAAGGAGAGAAACAACTTGAACCTTGT---TGCGCAT 1293
 Qy 1463 aacagacaacaacagaggggacgagcgaggaagagagagagagagagagagagagag 1522
 Db 1294 AAGAACAACAACAGAGCGCAGCAACAGGAAGAGACACCTTTGGAA----- 1338
 Qy 1523 ggaagtaacagagaggtgcgttaggtacacagcgaggttgaaggaagggcgatgttcac 1582
 Db 1339 -----GTGCGGAAATATAGAGCTGAATTTCTTGAACAAGATATATTGTA 1383
 Qy 1583 atgcacagctcatccagtagccatcaacgcttctccgaaactccatcctgctgtgcttc 1642
 Db 1384 ATCCACGAGGTTATCCAGTTGTTGGTCAACCTACCTCAGATCTGAATTTCTTGTCTTTT 1443
 Qy 1643 ggtatcaacgctgaaaaaaccacacagaatctccttgcaggtgataaagacaaatgata 1702
 Db 1444 GGTATCAATCCGAGAGAACACCAGAGGAACCTCTCTTCAGAGTTTGAAGACAAATGTGATA 1503
 Qy 1703 gaccagatagagaagaagaagagatttagcattccctcggtcggtgagcaagttgaag 1762
 Db 1504 AGCAGATACCTATCAAGTGCAGAGCTTGCCTTCCCTGGGTCTGCAAAAGATATTGAG 1563
 Qy 1763 aagctcatcaaaaaccacagaggaatctcaactttgtgagtgctgctcctca 1812
 Db 1564 AACCTAATAAGAGCCAAAGTGAGTCTCTACTTTGTGATGCTCAGCCTCA 1613

RESULT 14
 PSA276875
 LOCUS
 DEFINITION Pisum sativum mRNA for convicillin (cvc gene).
 ACCESSION AJ276875
 VERSION AJ276875.1 GI:7339550
 KEYWORDS convicillin; cvc gene; seed storage protein.
 SOURCE
 ORGANISM Pisum sativum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

ORGANISM *Vicia narbonensis*
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Viciaeae;
Vicia.

REFERENCE 1 (bases 1 to 1858)
 AUTHORS Koch, G., Koenig, S., Becker, C., Horstmann, C. and Schlesier, B.
 TITLE PCR cloning of cDNA for convicillin, a 7S globulin from *Vicia narbonensis* L
 JOURNAL unpublished
 REFERENCE 2 (bases 1 to 1858)
 AUTHORS Koch, G.
 TITLE Direct Submission
 JOURNAL Submitted (29-APR-1996) Koch G., Institut fuer Pflanzen-genetik und Kulturpflanzenforschung, Molekulare Zellbiologie, Corrensstr.3, Gatersleben, Sachsen-Anhalt, Germany, D-06466

FEATURES
 source Location/Qualifiers
 1..1858 /organism="Vicia narbonensis"
 /db_xref="taxon:3912"
 /clone="CV4"
 /tissue_type="cotyledon"
 /clone_lib="PCR-amplified cDNA"
 /dev_stage="seed, cell expansion stage"
 1..34 /codon_start=1
 35..118 /product="convicillin precursor"
 35..11672 /protein_id="CAA96513.1"
 5'UTR /db_xref="GI:1297070"
 sig_peptide /db_xref="SPTREMBL:O41674"
 CDS /translation="WATANKSRFPVLLGLLIIILASLCVTVYANDEGTEPRVQQRER
 GQEGEKRRKQWRPHEKQAQRRERWETSEERKVEDEWRGSRQHRDEPRER
 ERYABERKRWEGEKEGSKQSRERNPFLFKSNKFLTLFENENGHRRQRFDR
 RDLFENQYRLVYRAKRFIFLPHQIDADLILVLSGRAILTVLSPDRNSYNLE
 RGTIKPAGTTSYLLNDDDEDLVLDLSVNRPGKVSFGLSGSKNOYLRFKSN
 ILEASLNTYETIKVLLPEPOOSIGQRKRSQRTNALVRVSEQVEELKRLAKSS
 KGVSEPRFPNLRNPNKPKLFEITPEKKYPOLOLDLIFVSSVEINEGMLL
 PHNSRAIVLLVNSGKNLVLGKNSQQREREDQVORVEARLSPGDVLIIPA
 GHPVSAASNLLNLFGLGINENRNFLTGSDDNVIQSINPNVKELTFPGSAQEVNR
 LUKNQHSFHNLEPQEGESQRRKSPISLSLGFNF"
 119..1669
 mat_peptide /note="c DNA derived amino acid sequence of bases 119-142
 are confirmed by peptide sequencing"
 /function="seed storage protein"
 /product="7S globulin; convicillin"
 1673..1858
 3'UTR 660 a 305 c 414 g 479 t
 BASE COUNT
 ORIGIN

Query Match 17.1%; Score 347.8; DB 8; Length 1858;
 Best Local Similarity 59.2%; Pred. No. 2e-85;
 Matches 742; Conservative 0; Mismatches 442; Indels 59; Gaps 6;

Qy 561 ggaacaacctttactccctcaaggcggtttagcaccgcgctacgggaaccacaacg 620
 Db 423 GAAGAAACCCCTTCTCTCAAGTCTAATAAGTTTCTGACACTTTTCGAAAAATGAAACG 482

Qy 621 gtadgaccgggtcctcaagagtttgaccacaagggtcaaggcagtttcagaatctccaga 680
 Db 483 GTCACATTCGTCGCGCTTCAAGAGTTGACAAACGTTTTCAGACTTTATTTGAAATCTTCAA 542

Qy 681 atcaccgtattgtcagatcagagccaaacctaacactctgttcttcccaagcagctg 740
 Db 543 ACTATCGCTTGTGGATATAGAGCAAAACCTCACACCATCTTCCTTCTCTCAGCACAATAG 602

Qy 741 atgctgataaacctcttctgttatccagcaaggcgcaaccgctgaccgtagcaaatggca 800
 Db 603 ATGCTGACTTAAATCCTTACCGCTCCTCAGTGGGAGAGCAATTTGACAGTGTGAGTCCTTA 662

Qy 801 ataacagaagagcttttaactcttgacgagggccatgcactcagaatccccatccggtttca 860

Search completed: December 27, 2001, 13:04:39
 Job time: 13945 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 27, 2001, 13:15:28 ; Search time 486.97 Seconds
(without alignments)
3577.397 Million cell updates/sec

Title: US-09-715-036-5
Perfect score: 2032
Sequence: 1 aataatcatatattcgc.....cggttggtgcgtgtttcttc 2032

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 930521 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N_Geneseq_1101:*

- 1: /SIDS2/gcgdata/geneseq/geneseq/NA1980.DAT:*
- 2: /SIDS2/gcgdata/geneseq/geneseq/NA1981.DAT:*
- 3: /SIDS2/gcgdata/geneseq/geneseq/NA1982.DAT:*
- 4: /SIDS2/gcgdata/geneseq/geneseq/NA1983.DAT:*
- 5: /SIDS2/gcgdata/geneseq/geneseq/NA1984.DAT:*
- 6: /SIDS2/gcgdata/geneseq/geneseq/NA1985.DAT:*
- 7: /SIDS2/gcgdata/geneseq/geneseq/NA1986.DAT:*
- 8: /SIDS2/gcgdata/geneseq/geneseq/NA1987.DAT:*
- 9: /SIDS2/gcgdata/geneseq/geneseq/NA1988.DAT:*
- 10: /SIDS2/gcgdata/geneseq/geneseq/NA1989.DAT:*
- 11: /SIDS2/gcgdata/geneseq/geneseq/NA1990.DAT:*
- 12: /SIDS2/gcgdata/geneseq/geneseq/NA1991.DAT:*
- 13: /SIDS2/gcgdata/geneseq/geneseq/NA1992.DAT:*
- 14: /SIDS2/gcgdata/geneseq/geneseq/NA1993.DAT:*
- 15: /SIDS2/gcgdata/geneseq/geneseq/NA1994.DAT:*
- 16: /SIDS2/gcgdata/geneseq/geneseq/NA1995.DAT:*
- 17: /SIDS2/gcgdata/geneseq/geneseq/NA1996.DAT:*
- 18: /SIDS2/gcgdata/geneseq/geneseq/NA1997.DAT:*
- 19: /SIDS2/gcgdata/geneseq/geneseq/NA1998.DAT:*
- 20: /SIDS2/gcgdata/geneseq/geneseq/NA1999.DAT:*
- 21: /SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:*
- 22: /SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2032	100.0	2032	22	Peanut allergen Ara
2	2032	100.0	2041	22	DNA encoding anaph
3	2025.6	99.7	2032	18	Peanut allergen Ara
4	1930	95.0	1930	20	Peanut allergen, A
5	1752.4	86.1	1949	18	Peanut allergen Ar
6	1749.8	86.1	1952	22	Anaphylactic anti
7	406.8	20.0	1320	19	Coding sequence fo
8	404	19.9	1818	19	Coding sequence fo
9	324.2	16.0	1920	19	Coding sequence fo
10	160.2	7.9	2140	19	Macadamia integrif
11	157.4	7.7	2171	19	Macadamia integrif

12	152.6	7.5	2171	19	AAV42310	Macadamia integrif
13	138.8	6.8	1867	13	AAQ20377	Sequence of 67 kD
14	98.8	4.9	1924	20	AAV72243	G. max SPP2 cDNA.
C 15	53.2	2.6	936	22	AAF58252	Oligonucleotide D1
C 16	53.2	2.6	936	22	AAF58254	Oligonucleotide D1
C 17	53.2	2.6	936	22	AAF58257	Oligonucleotide D1
C 18	53.2	2.6	936	22	AAF58259	Oligonucleotide D2
C 19	53.2	2.6	936	22	AAF58262	Oligonucleotide D2
C 20	53.2	2.6	936	22	AAF58255	Oligonucleotide D1
C 21	49.6	2.4	10732	21	AAAI0594	Gene encoding a su
22	49.4	2.4	1068	22	AAAI0594	Human zmsel cDNA,
23	47.6	2.3	936	22	AAF58252	Oligonucleotide D1
24	47.6	2.3	936	22	AAF58254	Oligonucleotide D1
25	47.6	2.3	936	22	AAF58257	Oligonucleotide D1
26	47.6	2.3	936	22	AAF58259	Oligonucleotide D2
27	47.6	2.3	936	22	AAF58262	Oligonucleotide D2
28	47.6	2.3	938	22	AAF58255	Oligonucleotide D1
C 29	45.2	2.2	1823	21	AAC35016	Oligonucleotide D1
C 30	45.2	2.2	1823	21	AAC36549	Arabidopsis thalia
31	44	2.2	4000	18	AAI91902	Arabidopsis thalia
32	43.8	2.2	441	21	AAAI5042	Mannose-1-phosphat
33	42.4	2.1	726	22	AAI22214	Mouse secreted exp
34	42.4	2.1	726	22	AAI47511	Probe #12147 for g
35	42.4	2.1	726	22	AAI07913	Probe #16197 used
36	42.4	2.1	1951	22	AAI12976	Probe #7904 used t
37	42.4	2.1	1951	22	AAI34335	Probe #2909 for ge
38	42.4	2.1	1951	22	AAI02894	Probe #3021 used t
C 39	41.8	2.1	227	22	AAI22239	Probe #2885 used t
C 40	41.8	2.1	227	22	AAI47532	Probe #12172 for g
C 41	41.8	2.1	227	22	AAI07937	Probe #16218 used
C 42	41.8	2.1	494	22	AAI13001	Probe #7928 used t
C 43	41.8	2.1	494	22	AAI34356	Probe #2934 for ge
C 44	41.8	2.1	494	22	AAI02918	Probe #3042 used t
45	41.2	2.0	575	22	AAI23315	Probe #2909 used t
						Probe #13248 for g

ALIGNMENTS

RESULT	1
AAAF90339	AAAF90339 standard; cDNA; 2032 BP.
ID	AAF90339 standard; cDNA; 2032 BP.
XX	AAF90339;
AC	AAF90339;
XX	23-JUL-2001 (first entry)
DT	Peanut allergen Ara h1 P41B cDNA.
DE	Peanut; allergen; Ara h1 P41B; transgenic plant; allergy; ss.
KW	Peanut; allergen; Ara h1 P41B; transgenic plant; allergy; ss.
OS	Arachis hypogaea.
XX	Key
PH	Location/Qualifiers
FT	misc_feature
FT	676...930
FT	/tag= "PCR amplified fragment"
XX	WO200136621-A2.
PN	25-MAY-2001.
PD	20-NOV-2000; 2000WO-US31657.
XX	19-NOV-1999; 99US-0167255.
XX	(UYAL-) UNIV ALABAMA A & M.
PA	Dodo HW, Arntzen CJ, Konan KN, Viquez OM;
PI	WPI; 2001-355630/37.
DR	Producing transgenic peanut plants that produce allergen-free seeds,
XX	
PT	

PT useful in non-allergenic foods, by antisense or sense co-suppression of
PT allergen-encoding genes -

XX Claim 20; Fig 5; 72pp; English.

XX The present sequence is that of peanut allergen Ara h1 P41B cDNA.
CC A portion of this gene is homologous to the corresponding region
CC of the peanut allergen Ara h1 P17 gene. This region has been PCR
CC amplified, cloned in transformation vectors (pUC18 and pBI4434) in
CC sense and antisense orientations and used to down-regulate Ara h1
CC P41B and Ara h1 P17 allergens in peanut. This is an example of the
CC method of the invention, which relates to the production of a
CC peanut plant having reduced, or undetectable, allergenic protein
CC (AP) content in its seed. A peanut plant cell is transformed with
CC a DNA construct containing an antisense AP gene and/or sense AP
CC gene, or their fragments, regenerated to plants, and fertile
CC transgenic plants that produce seeds with reduced AP content are
CC identified. The seeds are useful for preparation of allergen-free
CC foods.

XX Sequence 2032 BP; 628 A; 473 C; 530 G; 401 T; 0 other;

Query Match 100.0%; Score 2032; DB 22; Length 2032;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2032; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	aataatcatatattcatcatcatctataataagtagtagcaggagcaaatgagaggag	60
DB	1	aataatcatatattcatcatcatctataataagtagtagcaggagcaaatgagaggag	60
QY	61	ggttctccactgtagtggtagggatctgtctctggtctgagttctgcaacgca	120
DB	61	ggttctccactgtagtggtagggatctgtctctggtctgagttctgcaacgca	120
QY	121	tgccaaagtcatcctaccagagaagaaacagagaaacccctgcgcagaggtgcctcca	180
DB	121	tgccaaagtcatcctaccagagaagaaacagagaaacccctgcgcagaggtgcctcca	180
QY	181	gagttgtcaacagggaacccggatgacttgagcaaaaggcattcgagctcgcgtgcccaa	240
DB	181	gagttgtcaacagggaacccggatgacttgagcaaaaggcattcgagctcgcgtgcccaa	240
QY	241	gctcgagtagatctcgtgtgtgtctatgatctctcaggagacactggcacaacca	300
DB	241	gctcgagtagatctcgtgtgtgtctatgatctctcaggagacactggcacaacca	300
QY	301	acgttccctccaggagcggacacgtggccgcaacccggagactcagatgatgaccg	360
DB	301	acgttccctccaggagcggacacgtggccgcaacccggagactcagatgatgaccg	360
QY	361	ccgtcaaccccgagagaggaagcagcagatggggaccagctggaccgagagcgtga	420
DB	361	ccgtcaaccccgagagaggaagcagcagatggggaccagctggaccgagagcgtga	420
QY	421	aagagaagaagacttgagacaaccaagaagattgagcgcaccaagtcatcagcagcc	480
DB	421	aagagaagaagacttgagacaaccaagaagattgagcgcaccaagtcatcagcagcc	480
QY	481	acggaataaaggcccgaggaagagaagagaacaagagtggggaacacacaggtagcca	540
DB	481	acggaataaaggcccgaggaagagaagagaacaagagtggggaacacacaggtagcca	540
QY	541	tgtgagggaagaacatctcggaacaacccctttctactctccctcgaagcgggttagcac	600
DB	541	tgtgagggaagaacatctcggaacaacccctttctactctccctcgaagcgggttagcac	600
QY	601	ccgctacgggaacaaacggtaggtccgggtctctcagaggtttgaccaaagggtcaag	660
DB	601	ccgctacgggaacaaacggtaggtccgggtctctcagaggtttgaccaaagggtcaag	660
QY	661	gcagtttcagaaatccagaaatccaccgtatgtgagatcgagggcccaacactcactct	720
DB	661	gcagtttcagaaatccagaaatccaccgtatgtgagatcgagggcccaacactcactct	720

DB	661	gcagtttcagaaatccagaaatccaccgtatgtgagatcgagggcccaacactcactct	720
QY	721	tgttcttccaaagcagctgatgctgataacatccttgttatccacgaaggcgcaagccac	780
DB	721	tgttcttccaaagcagctgatgctgataacatccttgttatccacgaaggcgcaagccac	780
QY	781	cgtgaccgtagcaaatggcaataacagaaagagctttaactcttgacgagggccatgcact	840
DB	781	cgtgaccgtagcaaatggcaataacagaaagagctttaactcttgacgagggccatgcact	840
QY	841	cagaatcccatccggtttcttctactatcttgaaccgcatgacaacacagaacctcag	900
DB	841	cagaatcccatccggtttcttctactatcttgaaccgcatgacaacacagaacctcag	900
QY	901	agtgcataaaatcccatccgcttaacacacccgcccagtttgagatttcttcccg	960
DB	901	agtgcataaaatcccatccgcttaacacacccgcccagtttgagatttcttcccg	960
QY	961	gagcagccgagacaatcatctacttgcagggttctcagcagggaatacgtttgaggccgc	1020
DB	961	gagcagccgagacaatcatctacttgcagggttctcagcagggaatacgtttgaggccgc	1020
QY	1021	cttcaatcggaattcaatgatacagagaggtgctgttagaagagaatcgaggaggtga	1080
DB	1021	cttcaatcggaattcaatgatacagagaggtgctgttagaagagaatcgaggaggtga	1080
QY	1081	gcaagagagagagggcagagggcgatgagtagtactcggagtagtgagaacaatgaagagt	1140
DB	1081	gcaagagagagagggcagagggcgatgagtagtactcggagtagtgagaacaatgaagagt	1140
QY	1141	gatagtcataaagtgtcaagagagacgcttggaagaacttactaagcacgctaaatccgtctc	1200
DB	1141	gatagtcataaagtgtcaagagagacgcttggaagaacttactaagcacgctaaatccgtctc	1200
QY	1201	aaagaaagggtccgaagaagagggagagatcaccaaccaactcaacttgagagaagcga	1260
DB	1201	aaagaaagggtccgaagaagagggagagatcaccaaccaactcaacttgagagaagcga	1260
QY	1261	gcccgatctttctaaacacttgggaagttatttggaggtagaagccagaagaacccc	1320
DB	1261	gcccgatctttctaaacacttgggaagttatttggaggtagaagccagaagaacccc	1320
QY	1321	ccagcttcagagacctgacatgatgtcacctgtgtagagatcaaaagagaggtttgat	1380
DB	1321	ccagcttcagagacctgacatgatgtcacctgtgtagagatcaaaagagaggtttgat	1380
QY	1381	gctcccaacttcaactcaaaaggccatgggttatcgtcgtcgtcaacaaaggaaactgaaa	1440
DB	1381	gctcccaacttcaactcaaaaggccatgggttatcgtcgtcgtcaacaaaggaaactgaaa	1440
QY	1441	ccttgaaactcgtggctgtaagaaagagacaacaagaggggagcgcggaagaagaga	1500
DB	1441	ccttgaaactcgtggctgtaagaaagagacaacaagaggggagcgcggaagaagaga	1500
QY	1501	ggacgaagaacgaagaagaggggaaagtaacacagagaggtgcgtaggtacacagcaggtt	1560
DB	1501	ggacgaagaacgaagaagaggggaaagtaacacagagaggtgcgtaggtacacagcaggtt	1560
QY	1561	gaaaggaaggcgatgtgttcatcatgcccagcagctcatccagtagccatcaacgcttctcc	1620
DB	1561	gaaaggaaggcgatgtgttcatcatgcccagcagctcatccagtagccatcaacgcttctcc	1620
QY	1621	cgaactccatctcgttggctcggtatcaacgctgaaacacacacagaaatcttcttgc	1680
DB	1621	cgaactccatctcgttggctcggtatcaacgctgaaacacacacagaaatcttcttgc	1680
QY	1681	aggtgataaaggaacaattgtatagaccagatagagaagcgaagcgaagatttagcatccc	1740
DB	1681	aggtgataaaggaacaattgtatagaccagatagagaagcgaagcgaagatttagcatccc	1740
QY	1741	tgggtcgggtgaaacaagttgagaagctcatcaaaaaaccagaaggaatctcaactttgtgag	1800
DB	1741	tgggtcgggtgaaacaagttgagaagctcatcaaaaaaccagaaggaatctcaactttgtgag	1800

Query Match 99.7%; Score 2025.6; DB 18; Length 2032;									
Best Local Similarity 99.8%; Pred. No. 0;									
Matches 2026; Conservative 0; Mismatches 4; Indels 0; Gaps 0;									
Qy	1	aataatcatatataatcatcatatataatgtagtagcagaggagcaatgagagggag	60						
Db	1	aataatcatatataatcatcatatataatgtagtagcagaggagcaatgagagggag	60						
Qy	61	ggtttctccactgatgctgtgttagggatccttgccttgcttcaagtttctgcaacgca	120						
Db	61	ggtttctccactgatgctgtgttagggatccttgccttgcttcaagtttctgcaacgca	120						
Qy	121	tgcaagtcatcaacttaccagaagaacacagagaacccctgcgcagaggtgctctcca	180						
Db	121	tgcaagtcatcaacttaccagaagaacacagagaacccctgcgcagaggtgctctcca	180						
Qy	181	gagttgtcaacaggaaacccgatgacttgaagcaaaaggccatgcgagttctcgtaacca	240						
Db	181	gagttgtcaacaggaaacccgatgacttgaagcaaaaggccatgcgagttctcgtaacca	240						
Qy	241	gctcgagtatgactcctctgttctgtctatgactcctcgaggacacactggcaccaccacca	300						
Db	241	gctcgagtatgactcctctgttctgtctatgactcctcgaggacacactggcaccaccacca	300						
Qy	301	acgttccctccaggagcgacacgtgcccgaacccggagactacgatgatgacacg	360						
Db	301	acgttccctccaggagcgacacgtgcccgaacccggagactacgatgatgacacg	360						
Qy	361	ccgtcaaccccgagagaggaagagcccgatgggacacagctggaccgagggagcgtaga	420						
Db	361	ccgtcaaccccgagagaggaagagcccgatgggacacagctggaccgagggagcgtaga	420						
Qy	421	aagagaagaagactcggagacaacccaagagaagattggaggcgaccaagttcatcagcagcc	480						
Db	421	aagagaagaagactcggagacaacccaagagaagattggaggcgaccaagttcatcagcagcc	480						
Qy	481	acggaataaaggccgaaggaagagaggaacaaagatggggagacacccaggttagcca	540						
Db	481	acggaataaaggccgaaggaagagaggaacaaagatggggagacacccaggttagcca	540						
Qy	541	tgtgagggaagaacacatctcggaaacaccccttctacttcccgtcaaggcgttagcac	600						
Db	541	tgtgagggaagaacacatctcggaaacaccccttctacttcccgtcaaggcgttagcac	600						
Qy	601	ccgctacggggaacaaacagtgtagtcccggtctcgcagaggtttgacaaagggtcaag	660						
Db	601	ccgctacggggaacaaacagtgtagtcccggtctcgcagaggtttgacaaagggtcaag	660						
Qy	661	gcagtttcagaatctccagaatcacccgtattgtgcagatcgaggcccaacctaacactct	720						
Db	661	gcagtttcagaatctccagaatcacccgtattgtgcagatcgaggcccaacctaacactct	720						
Qy	721	tgtttctcccaagcacgctgatgctgataacatccttctgttatccagcaaggccaagccac	780						
Db	721	tgtttctcccaagcacgctgatgctgataacatccttctgttatccagcaaggccaagccac	780						
Qy	781	cgtgacogtagcaaatggcaataacagaaagagctttaattcttgacagggccatgcaact	840						
Db	781	cgtgacogtagcaaatggcaataacagaaagagctttaattcttgacagggccatgcaact	840						
Qy	841	cagaatcccatccggtttcattctctacatcttgaaccgcgatgacaacagaaactcag	900						
Db	841	cagaatcccatccggtttcattctctacatcttgaaccgcgatgacaacagaaactcag	900						
Qy	901	agtagctaaatctccatgcgcgtttaacacacccggccagtttgaggtttcttcccgcg	960						
Db	901	agtagctaaatctccatgcgcgtttaacacacccggccagtttgaggtttcttcccgcg	960						
Qy	961	gagcagccgagaccaatcatctacttctgcagggttcagcaggaaatcagttgagggccgc	1020						
Db	961	gagcagccgagaccaatcatctacttctgcagggttcagcaggaaatcagttgagggccgc	1020						

Qy	1021	cttcaatgcggaatctcaatgaatcacgagggctgctgttagaagagaatgcaggggtga	1080
Db	1021	cttcaatgcggaatctcaatgaatcacgagggctgctgttagaagagaatgcaggggtga	1080
Qy	1081	gcaagagagagagggcgagagggctgtagtactcggagtagtgagaaacaatgaaggagt	1140
Db	1081	gcaagagagagagggcgagagggctgtagtactcggagtagtgagaaacaatgaaggagt	1140
Qy	1141	gatagtcaaatgctcaaaagagcacatttaagaacttactaagcacgcgtataatccgtctc	1200
Db	1141	gatagtcaaatgctcaaaagagcacatttaagaacttactaagcacgcgtataatccgtctc	1200
Qy	1201	aaagaagggtccgagagaagaggagatatcaccaacccaactcaacttgagagaaggoga	1260
Db	1201	aaagaagggtccgagagaagaggagatatcaccaacccaactcaacttgagagaaggoga	1260
Qy	1261	gccgactcttcttaacaactttgggaagttatttgaagtgaggtgaagccagacaagaaccc	1320
Db	1261	gccgactcttcttaacaactttgggaagttatttgaagtgaggtgaagccagacaagaaccc	1320
Qy	1321	ccagcttcaggaactggacatgatgctcacctgtgtagagatcaaaagaaggagctttgat	1380
Db	1321	ccagcttcaggaactggacatgatgctcacctgtgtagagatcaaaagaaggagctttgat	1380
Qy	1381	gctcccacacttcaactcaaaagccatggttatcgtcgtcgtaacaaagggaactggaaa	1440
Db	1381	gctcccacacttcaactcaaaagccatggttatcgtcgtcgtaacaaagggaactggaaa	1440
Qy	1441	ccttgaactcgtgctgtaagaanaagagcaacaacagaggggagcgcggaagaagagga	1500
Db	1441	ccttgaactcgtgctgtaagaanaagagcaacaacagaggggagcgcggaagaagagga	1500
Qy	1501	ggacgaagagaagaagagagggaggaagtaacagagaggtgcgttagtacacagcgaggtt	1560
Db	1501	ggacgaagagaagaagagagggaggaagtaacagagaggtgcgttagtacacagcgaggtt	1560
Qy	1561	gaaggaaggcgatgtgttcatcatgccagcagctcctcagtagccatcaacgcttctcc	1620
Db	1561	gaaggaaggcgatgtgttcatcatgccagcagctcctcagtagccatcaacgcttctcc	1620
Qy	1621	cgaactccatctgcttggttcggtatcaacgctgaaacacacacacagaaatcttctctgc	1680
Db	1621	cgaactccatctgcttggttcggtatcaacgctgaaacacacacacagaaatcttctctgc	1680
Qy	1681	aggtgataaagacaatgtgataccagatagagaagcaagcgaaaggttagctattcctcc	1740
Db	1681	aggtgataaagacaatgtgataccagatagagaagcaagcgaaaggttagctattcctcc	1740
Qy	1741	tgggtcgggtgaacaagttagaagctcatcaaaacccagagaagaatctcaactttgtgag	1800
Db	1741	tgggtcgggtgaacaagttagaagctcatcaaaacccagagaagaatctcaactttgtgag	1800
Qy	1801	tgcctgcctcaactctcaatctcgaatccgctcgtctcctgagaagaagctcctctgagaa	1860
Db	1801	tgcctgcctcaactctcaatctcgaatccgctcgtctcctgagaagaagctcctctgagaa	1860
Qy	1861	agaggatcaagagagagaaacccaaggagggaggggttccactccttcaattttgaaggc	1920
Db	1861	agaggatcaagagagagaaacccaaggagggaggggttccactccttcaattttgaaggc	1920
Qy	1921	ttttaactgagaatggaggaacttgttatgataataagataagcgtctttgttact	1980
Db	1921	ttttaactgagaatggaggaacttgttatgataataagataagcgtctttgttact	1980
Qy	1981	ctactatccaaaaacttatcaataaaataaaacggtttgtcgtttgtttctcc	2032
Db	1981	ctactatccaaaaacttatcaataaaataaaacggtttgtcgtttgtttctcc	2032

RESULT 4
AAZ06382
ID AAZ06382 standard; DNA; 1930 BP.
XX

```

AAZ06382;
XX 09-NOV-1999 (first entry)
XX Peanut allergen, Ara h 1.
XX allergy; immune response; transgenic; allergen; epitope;
KW Immunoglobulin E; Ig E; binding site; peanut; ds.
XX
XX Arachis hypogaea.
XX
XX Key Location/Qualifiers
XX CDS 50..1930
XX /*tag= a
XX /product= "Ara h 1"
XX
XX WO9938978-A1.
XX
XX 05-AUG-1999.
XX
XX 29-JAN-1999; 99WO-US02031.
XX
XX 27-AUG-1998; 98US-0141220.
XX 31-JAN-1998; 98US-0073283.
XX 13-FEB-1998; 98US-0074590.
XX 13-FEB-1998; 98US-0074624.
XX 13-FEB-1998; 98US-0074633.
XX
XX {SOSI/} SOSIN H.
XX {UYAR-} UNIV ARKANSAS.
XX {UYNY } UNIV NEW YORK MT SINAI SCHOOL MEDICINE.
XX
XX Bannon GA, Burks AW, Sampson HA, Sosin H;
XX WPI; 1999-479189/40.
XX P-PSDB; AAY15244.
XX
XX Modified allergen with reduced IgE binding, useful for treating e.g.
XX allergies
XX
XX Disclosure; Page 32-33; 46pp; English.
XX
XX This is the nucleotide sequence of the Ara h 1 protein from Arachis
XX hypogaea. The Ara h 1 protein has 23 IgE (Immunoglobulin E) binding
XX epitopes, four of which are immunodominant (AAY15247, AAY15249, AAY15250
XX and AAY15263).
XX By modifying the IgE binding sites the ability of the allergen to
XX provoke an immune response is downregulated. The epitopes of the IgE
XX binding sites can therefore be modified in genetically engineered plants
XX and animals to elicit less of an allergic response.
XX
XX Sequence 1930 BP; 595 A; 455 C; 514 G; 366 T; 0 other;

Query Match 95.0%; Score 1930; DB 20; Length 1930;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1930; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 aataacatataatcacaatcatcatataagtagtagcagagagcaatgagagag 60
DB 1 aataacatataatcacaatcatcatataagtagtagcagagagcaatgagagag 60
QY 61 ggtttccactgatgtgtgtctgaggatcctgtctcgttccagtttctgcaacgca 120
DB 61 ggtttccactgatgtgtgtctgaggatcctgtctcgttccagtttctgcaacgca 120
QY 121 tgcgaagtcatcaccttaccagaagaacagagaaacccctgcgccagaggtctcca 180
DB 121 tgcgaagtcatcaccttaccagaagaacagagaaacccctgcgccagaggtctcca 180
QY 181 gagttgtcaacagggaaccggatgacttgaagcaaaaggcatgcgagtctcgctcaccaa 240
DB 181 gagttgtcaacagggaaccggatgacttgaagcaaaaggcatgcgagtctcgctcaccaa 240

QY 241 gctcagatgatcctcgttgtgtctatgatcctcagaggacacactggcaccacca 300
DB 241 gctcagatgatcctcgttgtgtctatgatcctcagaggacacactggcaccacca 300
QY 301 agtttccctccagggagcgagacacgtggccgaaccccgagagactacgatgaccc 360
DB 301 agtttccctccagggagcgagacacgtggccgaaccccgagagactacgatgaccc 360
QY 361 ccgtcaacccgaagagagagagcgccgatgggaccagctggagccagggagcggtga 420
DB 361 ccgtcaacccgaagagagagagcgccgatgggaccagctggagccagggagcggtga 420
QY 421 aagagaagaagactggagacacacaaagagattggagcgaccagtcacagcagcc 480
DB 421 aagagaagaagactggagacacacaaagagattggagcgaccagtcacagcagcc 480
QY 481 acggaataaaggcccgagagagagagagagagagagagagagagagagagagagag 540
DB 481 acggaataaaggcccgagagagagagagagagagagagagagagagagagagagag 540
QY 541 tgtgaggaagaacacatctcggaaacaccccttctactcctccgtcaaggcggttagcac 600
DB 541 tgtgaggaagaacacatctcggaaacaccccttctactcctccgtcaaggcggttagcac 600
QY 601 ccgtacgggaacacaaacggtaggtaccgggtcctcagaggtttgaccaaagggtcaag 660
DB 601 ccgtacgggaacacaaacggtaggtaccgggtcctcagaggtttgaccaaagggtcaag 660
QY 661 gcagtttcagaatctccagaaatcacctgattgtcagatcgagcgcaaacctaacactct 720
DB 661 gcagtttcagaatctccagaaatcacctgattgtcagatcgagcgcaaacctaacactct 720
QY 721 tgttttcccaagcacgctgatgctgataacatccttcttatccagcaagggtcaagccac 780
DB 721 tgttttcccaagcacgctgatgctgataacatccttcttatccagcaagggtcaagccac 780
QY 781 cgtgacctagcaaatgcaataacagaaagagctttaactcttgacgagggcccatcact 840
DB 781 cgtgacctagcaaatgcaataacagaaagagctttaactcttgacgagggcccatcact 840
QY 841 cagaatcccatccggtttcatttctcatcttgaacgcgcacatgaacacagacacctcag 900
DB 841 cagaatcccatccggtttcatttctcatcttgaacgcgcacatgaacacagacacctcag 900
QY 901 agtagctaaatctccatgccgttaacacacccggcagtttgaggatttctccggc 960
DB 901 agtagctaaatctccatgccgttaacacacccggcagtttgaggatttctccggc 960
QY 961 gagcgccgagaccaatcatcctacttgcagggttccagcaggaatacgtttgaggccgc 1020
DB 961 gagcgccgagaccaatcatcctacttgcagggttccagcaggaatacgtttgaggccgc 1020
QY 1021 cttcaatcggaattcaatgagatcagaggggtcgtctttagaagagaatcgagggtga 1080
DB 1021 cttcaatcggaattcaatgagatcagaggggtcgtctttagaagagaatcgagggtga 1080
QY 1081 gcaagagagagagggcagagcgatgagtagctcggagtagtgagaacaatgaagagt 1140
DB 1081 gcaagagagagagggcagagcgatgagtagctcggagtagtgagaacaatgaagagt 1140
QY 1141 gatgtcaaaagtgtcaaaaggcagcttgagaacttactaagcacgctaaatccgtctc 1200
DB 1141 gatgtcaaaagtgtcaaaaggcagcttgagaacttactaagcacgctaaatccgtctc 1200
QY 1201 aagaagaagctccgaagaagagagagagatataccaaacccaactcaactgagagagcgca 1260
DB 1201 aagaagaagctccgaagaagagagagagagatataccaaacccaactcaactgagagagcgca 1260
QY 1261 gccgatcttttaacaaactttgggaagttatttgaggtgaagccagacagaagaaccc 1320
DB 1261 gccgatcttttaacaaactttgggaagttatttgaggtgaagccagacagaagaaccc 1320
```

```
Qy 1321 ccagcttcaggaacctgacatgatgctccactgtgtagagatcaagaagagcctttgat 1380
Db 1321 ccagcttcaggaacctgacatgatgctccactgtgtagagatcaagaagagcctttgat 1380
Qy 1381 gctccacacttcaactcaaaagccatggttactgctgctgcaacaagaaggaactggaaa 1440
Db 1381 gctccacacttcaactcaaaagccatggttactgctgctgcaacaagaaggaactggaaa 1440
Qy 1441 ccttgaactcgtgctgtaagaagaagcaacaacagagagcgcgcggaagaagaga 1500
Db 1441 ccttgaactcgtgctgtaagaagaagcaacaacagagagcgcgcggaagaagaga 1500
Qy 1501 ggacgaagacgaagaagaagaggaagtaacacagagaggtgcgtagtacacagcaggtt 1560
Db 1501 ggacgaagacgaagaagaagaggaagtaacacagagaggtgcgtagtacacagcaggtt 1560
Qy 1561 gaaggaaggcgatggttcactcatgccagcagctcactcagtagccatcaacgctctcc 1620
Db 1561 gaaggaaggcgatggttcactcatgccagcagctcactcagtagccatcaacgctctcc 1620
Qy 1621 cgaactccatctgcttggcttcggtatcaacgctgtaaaacacacacagaaatctcttgc 1680
Db 1621 cgaactccatctgcttggcttcggtatcaacgctgtaaaacacacacagaaatctcttgc 1680
Qy 1681 aggtgataaggaacaatgtatgacacagatagagaagcaagcgaagatttagcattccc 1740
Db 1681 aggtgataaggaacaatgtatgacacagatagagaagcgaagatttagcattccc 1740
Qy 1741 tgggtcgggtgaacaagtgtgagaagctcatcaaaacacacagagaatctcatttgtgag 1800
Db 1741 tgggtcgggtgaacaagtgtgagaagctcatcaaaacacacagagaatctcatttgtgag 1800
Qy 1801 tgcctcctcctaactcctaactcctcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtc 1860
Db 1801 tgcctcctcctaactcctaactcctcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtc 1860
Qy 1861 agaggaatcaagagaggaacacacacacagaggaaggtccactcttcaatttgaaggc 1920
Db 1861 agaggaatcaagagaggaacacacacacagaggaaggtccactcttcaatttgaaggc 1920
Qy 1921 tttaaactga 1930
Db 1921 tttaaactga 1930

RESULT 5
AAT76612
ID AAT76612 standard; cDNA to mRNA; 1949 BP.
XX
AC AAT76612;
XX
DT 29-DEC-1997 (first entry)
XX
DE Peanut allergen Ara hi cDNA clone p17.
XX
KW Peanut; seed storage protein; allergen; allergy; hypersensitivity;
KW vaccine; anaphylactic shock; immunotherapy; therapy;
KW monoclonal antibody; ELISA; analysis; Ara hi; ds.
XX
OS Arachis hypogaea strain Florunner.
XX
FH Key Location/Qualifiers
FT CDS 3..1847
FT sig_peptide /*tag= a
FT mat_peptide /*tag= b
FT polyA_signal 1918..1923
FT /*tag= d
XX
PN W09724139-A1.
XX
```

```
PD 10-JUL-1997.
XX
PF 23-SEP-1996; 96WO-US15222.
XX
PR 04-MAR-1996; 96US-0610424.
PR 29-DEC-1995; 95US-0009455.
XX
PA (UYAR-) UNIV ARKANSAS.
XX
PI Bannon GA, Burks AW, Cockrell G, Helm RM, Stanley JS;
XX
XX WPI; 1997-363453/33.
DR P-PSDB; AAW22149.
XX
PT Peanut allergens Ara hi and Ara hii - used for vaccination and in
PT two-site monoclonal antibody based ELISA
XX
PS Claim 31; Page 183-185; 354pp; English.
XX
CC This cDNA clone, designated p17, codes for the major peanut
CC allergen Ara hi (AAW22149), which has multiple IgE binding
CC epitopes (see AAW24165-87). It was amplified from peanut seed cDNA
CC using a primer (see AAT76616) based on an isolated Ara hi peptide
CC (see AAW24206). The sequence shows significant homology with the
CC vicilin family of seed storage proteins of other legumes. The
CC gene is capable of producing a protein product in prokaryotic
CC cells that is recognised by serum IgE from a large proportion of
CC individuals with peanut hypersensitivity. Ara hi and Ara hii (see
CC AAW24164) can be used to raise monoclonal antibodies which are used
CC in a specific two-site MAb ELISA for the detection of Ara hi or Ara
CC hii (claimed). IgE-binding Ara antigen epitopes may be used in
CC vaccines to protect against allergic reactions to peanut allergens,
CC e.g. anaphylactic shock.
XX
SQ Sequence 1949 BP; 599 A; 455 C; 517 G; 378 T; 0 other;

Query Match 86.2%; Score 1752.4; DB 18; Length 1949;
Best Local Similarity 95.0%; Pred. No. 0;
Matches 1894; Conservative 0; Mismatches 46; Indels 54; Gaps 6;

Qy 48 caatgagaggagggtttctccactgatgctgtgtgtagggatcttctcgtgggttcaag 107
Db 1 caatgagaggagggtttctccactgatgctgtgtgtagggatcttctcgtgggttcaag 60
Qy 108 ttctgcaacgcatgccaagtcatcacttaccagaagaacacagaaacccctgcgcc 167
Db 61 ttctgcaacgcatgccaagtcatcacttaccagaagaacacagaaacccctgcgcc 114
Qy 168 agagggtcctccagagttgtcaacaggaacccggatgacttgaagcaaaaggcatcgagt 227
Db 115 agagggtcctccagagttgtcaacaggaacccggatgacttgaagcaaaaggcatcgagt 174
Qy 228 ctgctgacacaaagctcgatgatcctcgtgtgtgtctatctatctcgtgagacacactg 287
Db 175 ctgctgacacaaagctcgatgatcctcgtgtgtgtctatctcgtgtgtctctatg-----acactg 222
Qy 288 gcaccaccaacacagttccctccaggaggagcggaacgtggtggtggtggtggtggtggtggt 347
Db 223 gcgcaccaacacacagttccctccaggaggagcggaacgtggtggtggtggtggtggtggtggt 282
Qy 348 acgatgatgaccccgctcaaccccggaagagagaaagagggccggtgggaccagctggac 407
Db 283 acgatgatgaccccgctcaaccccggaagagagaaagagggccggtgggaccagctggac 342
Qy 408 cgaggagcgtgaaagagaagagactggagacaaccaagaagaattgagggcgaccaa 467
Db 343 cgaggagcgtgaaagagaagagactggagacaaccaagaagaattgagggcgaccaa 402
Qy 468 gtcatcagcgccacggaaataaagcccgaaagagagagagagagagagagagagagagagagag 527
Db 403 gtcatcagcgccacggaaataaagcccgaaagagagagagagagagagagagagagagagagag 462
```


Db 1020 tgtaattccagcagcttatccattgctgcacagctacctcaaacctcaatttctctgc 1079
 Qy 1639 cttcggtatcaacctgaacacacacagatactcttccttcgagtgataagacaatgt 1698
 Db 1080 ttttgggtatcaatgctgagacacacagagaaactctcttcgagcgagagacaatgt 1139
 Qy 1699 gatagaccagatagagaagcaagcgaaggatttagcattctccctgggtcggtgacaagt 1758
 Db 1140 ggtgaaggcagatagaagacaagtgcaggagcttggttccctgggtcgcacaagatgt 1199
 Qy 1759 tgagaagctatcaaaaaacacagagaagaaatcactttgtgagtgctcgtctctca 1812
 Db 1200 tgagagctattaagaagcagaggaagaaatcactttgtgtgctcagcctca 1253

RESULT 8

ID AAV17562 standard; cDNA; 1818 BP.
 AC AAV17562;
 XX
 DT 10-JUN-1998 (first entry)
 XX
 DE Coding sequence for the alpha subunit of beta-conglycinin.
 XX
 KW Beta-conglycinin; soybean seed protein; transgenic plant;
 KW seed storage protein profile; ss.
 XX
 OS Glycine max.
 XX
 PN WO9747731-A2.
 XX
 PD 18-DEC-1997.
 XX
 PF 10-JUN-1997; 97WO-US09743.
 XX
 PR 14-JUN-1996; 96US-0019940.
 XX
 PA (DUPO) DU PONT DE NEMOURS & CO E. I.
 XX
 PI Fader GM, Kinney AJ;
 XX
 DR WPI; 1998-052298/05.
 XX
 PT Suppression of specific classes of soybean seed protein genes -
 PT useful to change seed storage protein profiles of transgenic plants
 XX
 PS Disclosure; Page 29-30; 58pp; English.

This sequence represents the coding sequence for the alpha subunit of the soybean seed protein beta-conglycinin. The method of the invention is for reducing the quantity of a soybean seed storage protein (A), such as beta-conglycinin, in soybeans. It comprises: (a) constructing a chimeric gene comprising: (i) a nucleic acid fragment encoding a promoter that is functional in the cells of soybean seeds; (ii) a nucleic acid fragment encoding all or a portion of (A) placed in sense or antisense orientation relative to the promoter of (i); and (iii) a transcriptional termination region; (b) creating a transgenic soybean cell by introducing into a soybean cell the chimeric gene of (a); and (c) growing the transgenic soybean cells of (b) under conditions that result in expression of the chimeric gene of (a); where the quantity of one or more members of a class of (A) subunits is reduced when compared to soybeans not containing the chimeric gene of (a). The method is used to construct transgenic soybean lines where the expression of genes encoding (A) are modulated to effect a change in seed storage protein profile of transgenic plants. Modification of the seed storage protein profile can result in the production of novel soy protein products with unique and valuable functional characteristics.

Sequence 1818 BP; 561 A; 424 C; 427 G; 386 T; 0 other;

Query Match 19.9%; Score 404; DB 19; Length 1818;
 Best Local Similarity 61.2%; Pred. No. 2.2e-109;
 Matches 765; Conservative 0; Mismatches 410; Indels 75; Gaps 4;
 Qy 563 aacaaaccttctacttccogtcaagcggttagcacccgtacaggggaacaaacaggt 622
 Db 577 agaaaccttctctcgtctcaacaggttcgaacctctctcaaaaccaaattggt 636
 Qy 623 agatccggctcctcgagaggtttgacaaaagtcgaagtcagattcagaaatctccagaa 682
 Db 637 cgattcgtctccagaggttcaacaaacgctcccaacaaattcagaatctccgagac 696
 Qy 683 oacgtatttgacagtcgagggccaaacctacaactctgtttctcccaagcacgtgat 742
 Db 697 taccgctatttgaggttcaactccaaacccacacacctctctctcccaaccatgctgac 756
 Qy 743 gctgataacatcctgttatccagcgaaggcgaagccacgctgacgctgacaaatggcaat 802
 Db 757 gctgattcactctacgttatcccttaacgggactgacctcttcttccttgcgacaaacgac 816
 Qy 803 aacagaaagagctttaattcttgacgaggccatgcactcagaaatcccatcccggtttcatt 862
 Db 817 gacagagactcctacagacttcaatctggtgagtcgctgagagtcctccctcaggaaccaca 876
 Qy 863 tctacatcttgaaacgcctgaacacacagaaacctcagagtagtagctaaaatctccatgcc 922
 Db 877 tactatgtgtcaacctgacaaacgaacaaatctcagattataacactcgcctacccatccc 936
 Qy 923 gttacaacacccggccagtttgaggtttcttcccgcgagcagcgagcgaacaaatccatcc 982
 Db 937 gttacaagcctggtagatttgaggtttcttctctatctagcactgaagctcaacaatcc 996
 Qy 983 tacttcagggcttcagcaggaatacgttgagggccgcttcaatgcggaattcaatgag 1042
 Db 997 tacttgcagagattcagcaggaacatttttagggctctcagataccaaatctcgaggag 1056
 Qy 1043 atacggagggctgctgttagaagaaatgcagagaggtgagcaagagagagagggcagag 1102
 Db 1057 ataaacaggttctgttttagtagaggaagggcagcagcagaggggagcagaggtg --- 1113
 Qy 1103 cgatggagtagctcgagtagtagagaaatgaaggaggcgagcagcttcttcaaaagtag 1162
 Db 1114 -----caagagagcgtgattgtggaaatctctcgaaggaa 1146
 Qy 1163 cacgttgaaacacttactaaagcgcgttaacgtctctcaaaagaaaggctccgaaagagag 1222
 Db 1147 cagattcgggcactgagcaaacgtgccaaatctagttcaaggaaaacattttcttgaa 1206
 Qy 1223 ggagatatacccaaccaatcaacttgagagaggcgagcgcgagcttcttcttaacaacttt 1282
 Db 1207 g-----ataaaccttttaacttgagagcgcgcgaccccatctactccaacagctt 1257
 Qy 1283 gggagttatttgaggtagcagacagaaagaaaccccgacttcagacctgacatg 1342
 Db 1258 ggcagttctttgagatcaccccgag---agaaaaaccccgagcttcggagcttgatc 1314
 Qy 1343 atgtcactctgttagagatcaaaagaggtttgatgtctcccaactcaactcaaaag 1402
 Db 1315 ttctcagattgtggatagaacgagggagctctcttctacacacactcaattcaaaag 1374
 Qy 1403 gccatggttatcgtcgtcgtcaacaaggaactggaacactgaaactgactgctgtgaaga 1462
 Db 1375 gcgtagtgcatactggcaatgaatgaaggagagtcgaaacattgaaactgttggcctcaaa 1434
 Qy 1463 aaagagcaacaacagaggggacgcggcggaagagagagagacgaacgaagaagagag 1522
 Db 1435 gaacaacaacagagcagcaacaggaagagcaaccttg----- 1473
 Qy 1523 ggaagtaacagagaggtgctgtaggttacacagcgaggttggaaggaagcgatgtgttcac 1582
 Db 1474 -----gaagtgcggaataatagagccgaattgtctgaaacagatatattgta 1521
 Qy 1583 atgccagagctcatccagtagccatcaacgcttctctccgaactccatctgctgtgcttc 1642

Db 1522 atccgcaggttccaggttggccaacgctacacctcaaatctgaatttttgcattt 1581
Qy 1643 ggtataacgctgaaacacacagaaatcttccctgcaggtgataaggaacaaatgtgata 1702
Db 1582 ggtattaatgccgagaaacacagagaaacttccctgcaggttcgaagacaatgtgata 1641
Qy 1703 gaccagatagagaagcaagcgaaggtatttagcattccctgggtcggtggaacaaagttgag 1762
Db 1642 agccadatacctagtcaagtgcagagctgtgattccctgggttcgcaagctgttggag 1701
Qy 1763 aagctcataaaaacacagagaaatctcactttgtgagtgctgcctcca 1812
Db 1702 aagctattaagaacaaagagaatcctactttgtgagtgctgcacgctaa 1751
RESULT 9
AAV17563
ID AAV17563 standard; cDNA; 1920 BP.
XX
AC AAV17563;
XX
DT 10-JUN-1998 (first entry)
XX
DE Coding sequence for the alpha' subunit of beta-conglycinin.
XX
KW Beta-conglycinin; soybean seed protein; transgenic plant;
KW seed storage protein profile; ss.
XX
OS Glycine max.
XX
PN WO9747731-A2.
XX
PD 18-DEC-1997.
XX
PF 10-JUN-1997; 97WO-US09743.
XX
PR 14-JUN-1996; 96US-0019940.
XX
PA (DUPO) DU PONT DE NEMOURS & CO E. I.
XX
PI Fader GM, Kinney AJ;
XX
XX WPI; 1998-052298/05.
DR
XX Suppression of specific classes of soybean seed protein genes -
PT useful to change seed storage protein profiles of transgenic plants
XX
PS Disclosure; Page 30-31; 58pp; English.
XX
CC This sequence represents the coding sequence for the alpha' subunit of
CC the soybean seed protein beta-conglycinin. The method of the invention is
CC for reducing the quantity of a soybean seed storage protein (A), such as
CC beta-conglycinin, in soybeans. It comprises: (a) constructing a chimeric
CC gene comprising: (i) a nucleic acid fragment encoding a promoter that is
CC functional in the cells of soybean seeds; (ii) a nucleic acid fragment
CC encoding all or a portion of (A) placed in sense or antisense orientation
CC relative to the promoter of (i); and (iii) a transcriptional termination
CC region; (b) creating a transgenic soybean cell by introducing into a
CC soybean cell the chimeric gene of (a); and (c) growing the transgenic
CC soybean cells of (b) under conditions that result in expression of the
CC chimeric gene of (a); where the quantity of one or more members of a
CC class of (A) subunits is reduced when compared to soybeans not containing
CC the chimeric gene of (a). The method is used to construct transgenic
CC soybean lines where the expression of genes encoding (A) are modulated to
CC effect a change in seed storage protein profile of transgenic plants.
CC Modification of the seed storage protein profile can result in the
CC production of novel soy protein products with unique and valuable
CC functional characteristics.
XX
XX Sequence 1920 BP; 634 A; 444 C; 449 G; 393 T; 0 other;

Query Match 16.0%; Score 324.2; DB 19; Length 1920;
Best Local Similarity 58.5%; Pred. No. 1.2e-85;
Matches 764; Conservative 0; Mismatches 408; Indels 135; Gaps 6;
Qy 563 acaaacctttctactccgtcaagggcggttttagcaccgctacggggaacccaacgggt 622
Db 625 agaaacctttctactccgtcaagggcggttttagcaccgctacgttcccaaaaacatacgc 684
Qy 623 agaatccgggttctcagaggtttgaccaaaggtcaagggcagtttcagaatctccagaat 682
Db 685 caggttcgggttctcagaggtttcaaaaacgctcccaacaggttcagaatctccgagac 744
Qy 683 caccgtattgtcagatcgaggccaaacaaactaactctgttcttcccaagcagctgat 742
Db 745 taccgattttgttgagttcaactcccaaacacctctctctccccaccacctgcgcac 804
Qy 743 gctgataacatctctgttatccagcaagggcaagccacgctgacccgtgacagtaaggcaat 802
Db 805 gctgattacctctctctctctctcttaacgggactgcattcttactctgttgaaacaagac 864
Qy 803 aacagaagaagctttaatttgacgagggccatgcactcagaatccc----- 849
Db 865 gaccgagactcttacaaccttcaatctgcgctgacgtccctgaagctccctgaggaaccaca 924
Qy 850 -----atccggtttcattcc 865
Db 925 ttctatgtgtttaaccttgacaacgacgagaaatctcagaatgatagcaggaaaccacattc 984
Qy 866 tacatcttgaaacgctcagatgacaacccacgactcagagtagctaaaaatctcactccggtt 925
Db 985 tatgtgttaaccttgacaacgacgagaaatctcagaaatgaatacactcgccataccggtt 1044
Qy 926 aacacacccgctcaggtttgagattcttcccgcgagcgcgagaccagaccatcatctac 985
Db 1045 acaaacccggttagatttgagagttctctctctatctagcactcaagctcaacagctctac 1104
Qy 986 ttgcagggcttcagcaggaatacgtttgagggccgcttcaatcgcggaattcaatgagata 1045
Db 1105 ttgcaagggttcagcaagaatattctagaggctcatacagcaccacaaatctcgagagata 1164
Qy 1046 cggagggctgtttagaagagaatcagaggtgagcaggaagagagagggcagagcgca 1105
Db 1165 aacaagggtctgtgttgtagagggagggcgagcaacaagggggagaggtgc----- 1219
Qy 1106 tggagtactcggagtagtgagaaatgaaggagtgatgtaaaagtgtcaaaagtgtcaaaagagcac 1165
Db 1220 -----aagagagtggtgattgtggaatctcaagaacaa 1254
Qy 1166 gttgaagaacttactaagcagcgttaaatccgtctcaaaagaaggctccggaagagagga 1225
Db 1255 attcgggaactgagcaaacatgcgaatctagttcaaggaaaccattttcttgaag-- 1312
Qy 1226 gatacaccaaccccaatcaacttgagagagggcagcgcgctcttcttaacaactttggg 1285
Db 1313 -----ataaaccttcaacttgggaagcgcgacccactctattccaacaagcttggc 1365
Qy 1286 agttatttgagtgagccagacagaagaagacccccagcttcagggacctggacatgatg 1345
Db 1366 aagttgttgagattacc-----agagaacacctcagcttcgggacttggatgtctc 1419
Qy 1346 ctcaactggttagatcaaaagaaggagcttctgtctccacacttcaactcaaaagccc 1405
Db 1420 ctcaagtggtgatatgaagagggagctcttttctacacactcaattcaaaagccc 1479
Qy 1406 atggttatcgtcgtcgtcaacaagaaggaactggaaaccttgaaactcgtggtgtaagaaa 1465
Db 1480 atagtggtactagtattgaatgaagagagaagcaaacattgaacttgt---tgccattaaa 1536
Qy 1466 gagcaacaacagagggcgcggaagaagagagagagagagagagagagagagagagagag 1525
Db 1537 gaacaacaacagagagcagcaacaggaagagcaacctttggaa----- 1578
Qy 1526 agtaacagagaggtgcgttaggttacacagcagaggttggaagggcgatgtgttccatcgt 1585

```
Db 1579 -----gtgcgaataatagactgaattgtctgacaagaatatattgtaatc 1626
Qy 1586 ccagcagctcatcagtagcatcaacgcttctccgaactccatctgcttgctcggt 1645
Db 1627 ccagcaggttatccagttatggtcaacgctacctcagatctgaattcttctgttgg 1686
Qy 1646 atcaacgtgaaacaacacagaaatctcttcgtaggtgataaggacaaatgtgatagc 1705
Db 1687 atcaatgccgagaaacaacagaaatctcttcgtaggttcgaaagacaatgtgataagc 1746
Qy 1706 cagatagagaagcaagcaaggtatttagcattccctgggtcggtgaaacaattgagaag 1765
Db 1747 cagatacctatgcaagtgcagggagcttgcttccctaggtctgcaaaagatatgagaac 1806
Qy 1766 ctcatcaaaacacagaaaggaatctcacttctgtgagtgctgctctca 1812
Db 1807 ctaataaagagcaaaagtgaagtactacttctgtgagtgctgagctctca 1853

RESULT 10
AAV42316
ID AAV42316 standard; cDNA; 2140 BP.
XX AC AAV42316;
XX DT 27-OCT-1998 (first entry)
XX DE Macadamia integrifolia partialantimicrobial protein gene.
XX KW antimicrobial protein; infestation; control; ss.
XX OS Macadamia integrifolia.
XX FH Key Location/Qualifiers
XX FT CDS 1..1878
XX FT /*tag= a
XX FT /product= partial antimicrobial protein
XX FT mat_peptide 1..1875
XX FT /*tag= b
XX PN WO9827805-A1.
XX PD 02-JUL-1998.
XX PF 22-DEC-1997; 97WO-AU00874.
XX PR 20-DEC-1996; 96AU-0004275.
XX PA (RETR-) COOP RES CENT TROPICAL PLANT PATHOLOGY.
XX PI Bower NI, Goulter KC, Green JL, Manners JM, Marcus JP;
XX WPI: 1998-377279/32.
XX DR P-PSDB; RAN62830.
XX Novel anti-microbial protein from e.g. Macadamia integrifolia -
PT useful for controlling microbial infestations of plants or mammals
XX Claim 5; Page 46-47; 96pp; English.
XX The sequence is that encoding an antimicrobial protein which can
CC be used to control microbial infestations in plants and mammalian
CC animals.
XX Sequence 2140 BP; 695 A; 486 C; 557 G; 402 T; 0 other;
SQ

Query Match 7.98; Score 160.2; DB 19; Length 2140;
Best Local Similarity 48.5%; Pred. No. 6.4e-37;
Matches 705; Conservative 0; Mismatches 658; Indels 90; Gaps 6;
Qy 415 gcgtgaaagagaagactggagacaacaaagagaagattggaggcgaccaaagtcata 474
```

```
Db 471 gcggcactggaaacaacagagagccagctctcgagtaccagtgccaggaagtccaaga 530
Qy 475 gcagccacggaaataaagcccgaagaaagagagagaaagagagagagagagagagagag 534
Db 531 gcagcagaggcaacacgcccagagtgctgatttgatgaacccctcagagggagggagcgg 590
Qy 535 tag---ccatgtgagggagaaacacatctcggaaacaccccttctactctccgtcaaggcg 591
Db 591 cagatacagagagggagagagagaaagcaaaagcgacaacccctactactctcagaaagag 650
Qy 592 gtttagcaccctgaaggggaacccaaacggtaggatcccggtgctcgagagaggttgacca 651
Db 651 ctttaagtacaaggttcagggaccgaggaagcccacatctcagttcttggaagacttatgg 710
Qy 652 aagtcgaagcagtttcagaatctccagaatccagctatttgtagagatcgagcgaacaaac 711
Db 711 tagatccaagcttctacgcgcaactaaaaactatcgttgggtcctcagggctaaaccc 770
Qy 712 taacactctgttctcccaagcacgctgatgctgataacatcccttgttatccagcaagg 771
Db 771 caagccttctgctcctaccacttggatgcagatgccattctcttctgttgcacggagg 830
Qy 772 gcaagccaccgtgaccgtagcaaatggcaataacagaaagagcttttaattcttgagcaggg 831
Db 831 gaggaggccctcaaaatgatccacggtgacaacagagaaatccctacaacctcgatgtgg 890
Qy 832 ccactgactcagaatcccatccggttttcttctacatttgaacccgcatgacaacca 891
Db 891 agcgttaatacagaatcccgctggaaacacattctacttaatacaaccgagacaacaaga 950
Qy 892 gaacctcagagtagctaaataatccatcccgcttaacacacccgcccagtttgagagatt 951
Db 951 gaggctccatagccaagttcttcagaccataccatccctcctggccatacaaggaaatt 1010
Qy 952 ctcccggcgagcgcgagaccaatccatctacttgcagggttcagcaggaaatcacgtt 1011
Db 1011 ctcccgagctggaggccaaacccagagcgcgtacacctcagtaacctcagcaagagattct 1070
Qy 1012 ggaggccgcttcaatgcggaattcaatgatacagagaggggtgctgtttagaagagaatgc 1071
Db 1071 cgaggctgcgtcaacacacaacagagagggctgcgtgggggtct----- 1115
Qy 1072 aggaggtgagcaagagagagagggcagagcgatggagtactcgagtagtgagaacaa 1131
Db 1116 -----tggaacagcaag 1127
Qy 1132 tgaaggagtagatgcaaaagtgtcaaaaggagacgcttgaaagaacttact---aagcacgc 1188
Db 1128 ggaggagtagataaattaggcgtcaacagggagcagacagggaggtgactcgagatgactc 1187
Qy 1189 taatccgctctcaaaagggctccgaaagagggagagatatacccaacccaactcaactt 1248
Db 1188 agagtccagcgtggtcatataaggagaggtggtgaatcaagcaggggaccttacaatct 1247
Qy 1249 gagagaagggcagcccgatcttctcaacacttgggaaggtatttgaggtgaagccaga 1308
Db 1248 gttcaacaaaggccactgtactcccaacaaatcgcgtcaagcctcgaagtgcaaacctg 1306
Qy 1309 caagaagaaccccgacttcaggacctggacatgatctcacctgtgtagagatacaaga 1368
Db 1307 --agactacagccaactccaagacatggacgtatcagtttctatagccaacatcaccca 1364
Qy 1369 agggacttgatgctccacacttcaactcaaaaggccatgggttatcgtcgtcgtcgaacaa 1428
Db 1365 gggatccatgatgggtccctcttcaacactaggtctacaaggtggtgtagtggctag 1424
Qy 1429 aggaactggaaacctgaactcgtggctgtaagaaaaagagacaacagaggggagcgcg 1488
Db 1425 tgggagggcagatcggaatcgcaatcgccctcactctgtcgggaagacacggcgccgcgg 1484
Qy 1489 ggaagagagagagacgaacgaagagagggagggaagtaacagagagaggtgcgtaggtta 1548
```

Db 1485 tgagggaaagcatgagaggaagagaggtgcactatgagcagg-----t 1532

QY 1549 cacagcaggttgaagaaagcgatgtgttcacatgcacgagctcactagccat 1608

Db 1533 tagagcagttgtcgaagagagagcattgttttgcaggtcactccgtgttt 1592

QY 1609 caacgcttccctgcgaactccatctgc-----ttggttcggtatcaacgctgaaacaa 1662

Db 1593 cgtttcatcccgaaacgaaacactgctgtttttgcatgtgaatcaatgccaaacaa 1652

QY 1663 ccacagaatcttccgtgcaggtgataagacaaatgtatagaccagatagagaagcagc 1722

Db 1653 ccacgagaaacttccgtgcgagagagaggaacgtgtgcagcagatagagccacagc 1712

QY 1723 gaaggattagcttccctccgtgcggtgaaacaggttgagaagctcatcaaaacccagaa 1782

Db 1713 aatggagctagcgttgcgccttcaagaaagagagtagaagagttatttaacagccagga 1772

QY 1783 ggaatctcaattgtgagtgctgctctcaatctcaatctcaatctcgtgtctctga 1842

Db 1773 cgagtctatcttcttccgtggccagcagcagcagcagcagcagcagcagcagcagcagc 1832

QY 1843 gaagaggtctcct 1855

Db 1833 gcaacacagcct 1845

RESULT 11

AAV42311

ID AAV42311 standard; cDNA; 2171 BP.

AC AAV42311;

XX 27-OCT-1998 (first entry)

DE Macadamia integrifolia antimicrobial protein gene.

XX antimicrobial protein; infestation; control; ss.

XX Macadamia integrifolia.

XX Key Location/Qualifiers

FT CDS 1..2001

FT /*tag= a

FT /product= antimicrobial protein

FT sig_peptide 1..86

FT /*tag= b

FT mat_peptide 87..1999

FT /*tag= c

XX WO9827805-A1.

XX 02-JUL-1998.

XX 22-DEC-1997; 97WO-AU00874.

XX 20-DEC-1996; 96AU-0004275.

XX (RETR-) COOP RES CENT TROPICAL PLANT PATHOLOGY.

XX Bower NI, Goulter KC, Green JL, Manners JM, Marcus JP;

XX WPI; 1998-377279/32.

XX P-PSDB; AAW62829.

XX Novel anti-microbial protein from e.g. Macadamia integrifolia -

XX useful for controlling microbial infestations of plants or mammals

XX Claim 5; Page 41-43; 96pp; English.

XX The sequence is that encoding an antimicrobial protein which can

XX be used to control microbial infestations in plants and mammalian

XX animals.

XX SQ Sequence 2171 BP; 680 A; 509 C; 571 G; 411 T; 0 other;

Query Match 7.7%; Score 157.4; DB 19; Length 2171;

Best Local Similarity 48.9%; Pred. No. 4.4e-36;

Matches 649; Conservative 0; Mismatches 591; Indels 87; Gaps 5;

QY 538 ccatgtgagggagaaacatctcgaacaccccttctacttccgtcaagcggtttag 597

Db 720 cgaggaggagagagaaagcgaacacccctacttctacttgcagaaagccttaag 779

QY 598 caccgctacggggaacacaaacggttaggtcgggtcgtcagaggtttgaccaaagtc 657

Db 780 tacaaggttcaggaacgaggaagccacatctcagttcttggaacttctatgtagatc 839

QY 658 aaggcagtttcagaatctccagaatccagatctgtgcagatcgagccaaacctaacc 717

Db 840 caagcttcacgcgcactaaaaactatcgttgggtcgtcctcaggttaaccccaacgc 899

QY 718 ttgtgttctcccaagcagctgctgataaacatcttgcacagcaagggcagc 777

Db 900 cttcgtgtcctaccacttggacgagatgccattcttggtcaccggagggaggg 959

QY 778 caccgtgacgttagcaaatgccaataacagaaagagctttaattgtgacagggccatgc 837

Db 960 agccctcaaaatgatccaccgtgacaacagagaaatcctacacacctgagtgtagacgt 1019

QY 838 actcagaatcccatcgggttctacttctacatcttgaaacccgcatgacacagaaacct 897

Db 1020 aatcagaatcccgctggaaacacatctacttaataccacggagacacacagagagct 1079

QY 898 cagagttagtaaaatctccatgcctgttaacacacccggccagtttgagatttctccc 957

Db 1080 ccacatagccaagttcttacagaccataccactcctggccaatacaagaaattctccc 1139

QY 958 ggcgagcagccgagacaaatcactctacttgcaggtctcagaggaatactgttgaggc 1017

Db 1140 agctggagggccaaacccagagcgtactcctcagtcacacgagaaagattctcagggc 1199

QY 1018 cgcttcaatgcgggaattcaatgagatcagggaggtgctgttagaagaagatgcaggag 1077

Db 1200 tgcgtcaacacacagcagagaggtgctggtgggtgct----- 1238

QY 1078 tgagcaagggagagggcagagggcagatgagtagctcaggtagtagagacaatgaagg 1137

Db 1239 -----tggaacagcaaaagggagg 1256

QY 1138 agtgatagtcgaagtcaaaaggagcagcttgaaagaacttact---aagcacgctaaatc 1194

Db 1257 agtgataattagtcggtcacagggagcagatcagggagttgactcgagatgactcagatc 1316

QY 1195 cgtctcaaaagaaggctccgaagaagagggagatatcaccaccccaatcaacttgagaga 1254

Db 1317 acgacgtggcataaaaggagaggtgtgaaatcaagcagggggaccttacaatctgttcaa 1376

QY 1255 aggcagcccgatcttctcaacactttgggaagttattgagttgaagccacacagaa 1314

Db 1377 caaaagggcactgtactccaaataacggtcaagcctacgaagctcaaacctg---agga 1433

QY 1315 gaaccccgagcttcaggacctggacatgatgctcacctgtgtagagatcaaaagaagagc 1374

Db 1434 ctacaggcaactccaagacatggcgtatcgttttctcatagccaacatcaccagggatc 1493

QY 1375 ttgtatgtcccaacttcaactcaaaagggcattggttatcgtcgtcgtcacaagaagac 1434

Db 1494 catgatgggtccctcttcaactactagttctcaaaaggtgtagtgggtgtagtagaga 1553

QY 1435 tggaaaccttgaactcgtgctgtgaaagaaagagcaacacagagggggcgccggggaaga 1494

Db 1554 ggcagatgtggaatggcatgccctcacttgcgggaagacacgcgcgcctggagg 1613

QY 1495 agaggagacgaagacgaagaggagggaagtaacagagaggtgcgtagggtacacagc 1554

Db 1614 gaaagcatgaggaggaagatgtgcactatgag-----caggttaaacg 1661
 QY 1555 gaggttgaagaagcgatgtgttccatcagccagcagctccatccagtagccatcaacgc 1614
 Db 1662 acgtttgtcgaagagagagccattgtttccggttaggtccatcccgctcgtctcttcc 1721
 QY 1615 ttc-----ctccgaactccatcgttgcgttcgtatcaacgctgaacacacacag 1668
 Db 1722 atccgaaacagaaacctgctgttttgcatttggaaatcaatgccaaacacacacga 1781
 QY 1669 aatcttctcagtgatagaaacaaatgtgatagaccagatagagaagcaagcgaagga 1728
 Db 1782 gaacttctcgcgggagagagagaaacgtgtcagcagatagagccacaggaatgga 1841
 QY 1729 tttagcattccctgggtcgggtgaacaaagttgagaagctcatcaaaacacagaagaatc 1788
 Db 1842 gctagcttgcctccaaagaaagaggttagaagagttatttaacagccagacagtc 1901
 QY 1789 tcactttgtagtgctcctcaatctcaatctcaatctcaatctcctcgtctcctgagaaaga 1848
 Db 1902 tatcttcttctcgtcccgccagcagccagcaacagcttcccgctcccaacgaacaa 1961
 QY 1849 gtctcct 1855
 Db 1962 acagcct 1968

RESULT 12

AAV42310

ID AAV42310 standard; cDNA; 2171 BP.

XX AC AAV42310;

XX 27-OCT-1998 (first entry)

XX DE Macadamia integrifolia antimicrobial protein gene.
 XX DE
 XX KW antimicrobial protein; infestation; control; ss.
 OS Macadamia integrifolia.

XX Key Location/Qualifiers

FH 1..2001

FT CDS /*tag= a

FT /*product= antimicrobial protein

FT sig_peptide 1..85

FT /*tag= b

FT mat_peptide 86..1999

FT /*tag= c

FT W09827805-A1.

XX PN 02-JUL-1998.

XX PD 22-DEC-1997; 97WO-A000874.

XX PF 20-DEC-1996; 96AU-0004275.

XX PR (RETR-) COOP RES CENT TROPICAL PLANT PATHOLOGY.

XX PA Bower NI, Goulter KC, Green JL, Mannes JM, Marcus JP;

XX PI WPI; 1998-377279/32.

XX DR P-PSDB; AAW62828.

XX XX Novel anti-microbial protein from e.g. Macadamia integrifolia -

XX PT useful for controlling microbial infestations of plants or mammals

XX PS Claim 5; Page 37-38; 96pp; English.

XX CC The sequence is that encoding an antimicrobial protein which can

XX be used to control microbial infestations in plants and mammalian

CC animals.
 XX Sequence 2171 BP; 687 A; 507 C; 570 G; 407 T; 0 other;
 SQ

Query Match 7.5%; Score 152.6; DB 19; Length 2171;

Best Local Similarity 48.7%; Pred. No. 1.2e-34;

Matches 646; Conservative 0; Mismatches 594; Indels 87; Gaps 5;

QY 538 ccatgtgagggaagaacatctcggaaacacccctttctactccgtcaaggcggttag 597
 Db 720 cgaggaggagaagagcaaacgcacccctactacttcgcagaaacgaaagcttaag 779
 QY 598 caccgcctacgggaacaaacacggtaggtccgggtcctcagaggtttgacaaaggtc 657
 Db 780 tacaaggttcagaccgaggaagccacatcctcagttcttgagaactctatgttagtc 839
 QY 658 aaggcagtttcagaaatctccgaatccacgtattgtgcagatcgagggccaaacctaacac 717
 Db 840 caagcttctacgcgcactaaaaaactatcgttgggtcctcgcaggttaaccccaacgc 899
 QY 718 tctgttctcccaagcacgcgtgatgctgataacatccttcttatccagcaaggccaagc 777
 Db 900 cttcgtcctccacacttgatgcagatgcacattctcttctgtcatagaggagagg 959
 QY 778 caccgtgaccgtagcaaaatggcaataacagaagaaggtttaattcttgacggggccatgc 837
 Db 960 agcctcaaaatgatccaccacgacacagagaatctacacactcgagtgtgagacgt 1019
 QY 838 actcagaatcccatccggtttcattctcctacatcttgaaaccgcacatgacaacacgaacct 897
 Db 1020 aatcagaatcccgctggaaacacattctacttaataccacggagacaacacagaggct 1079
 QY 898 cagagtagctaaaaatctccatcccgcttaacacaccccgccagtttgaggtattcttccc 957
 Db 1080 ccacatagccaagttcttacagaccatctccactcctcgtgccataacagaattctccc 1139
 QY 958 ggcgagacgcgagacccaatctcctacttgagggttcagcaggaatacgttggagc 1017
 Db 1140 agctggaggccaaacacagagccgctcctcagtcactctcagcaagagattctcagc 1199
 QY 1018 cgccttcaatcggaaattcaatgagatcacggaggtgctgtgtagaagagaatcagagag 1077
 Db 1200 tgcgctcaacacacacagagaagctgctgggtgtt----- 1238
 QY 1078 tgagcaagaggagagaggcgagagcgatggagtactcggagtagtgagaacaatgaag 1137
 Db 1239 -----tgacagcaagggagg 1256
 QY 1138 agtgatagtcaaaagttcaaaaggagcagcttgaaagaacttact---aagcacgctaaatc 1194
 Db 1257 agtgataattaggcgctcacagagcagatcaggaggttgactcgagatgactcagagtc 1316
 QY 1195 cgtctcaagaagggctccgagaagaggagagatatcaccaccccaatcaacttgagaga 1254
 Db 1317 acgacactggcatataagagaggtgtgaaatcaagcaggggaccttacaatgttcaa 1376
 QY 1255 aggcgagccgctcttctcaacaactttgggaagttatttgaggtgaagccagacaagaa 1314
 Db 1377 caaaggccactgtactccaacaaatcaggtcgaagcctacgaagtcgaacctg---agga 1433
 QY 1315 gaaccccccagcttcaggacctggacatgactcaccctgtgtagagatcaaaagagagc 1374
 Db 1434 ctacaggcaactccaagacatggacttatcggtttctatagccaacgtcccccaggtatc 1493
 QY 1375 tttagtgcctccacactcaactcaaaaggccatggttatcgtcgtctcaacaagaac 1434
 Db 1494 catgtggttccctctccacactaggtctcaaaaggtgtgtgtgtgtgtgtgtgtgtgt 1553
 QY 1435 tggaaaccttgaaactcgtggctgtaagaaaagagacaacacagagggggcgcggaaga 1494
 Db 1554 ggcagatgtgaaatggcatgccctcacttctcgggaagacacgcggcggtgtgagg 1613

Db 1248 gatattgggtacgtctcaaatgcttgccgcgcattctctccagacagagccaggatcccaaa 1307
Qy 1479 gggacgcggggaagaagagagacaaacacgaagaagaaggaaggaagtaacagagagg 1538
Db 1308 gtcggaagcagacagaagaacaaagaagaagagtcagaagaggagacatttgggagaat 1367
Qy 1539 tgcgtaggtacacagcagcaggttggaagaaagcgcgtgttccatcatgccacagcgtcatc 1598
Db 1368 tccagcaggtcaagaccatttccactgtgtacgtcttgcagcccgccagcgcacg 1427
Qy 1599 cagtagccatcaacgcttctccgaactccatctg-----cttgcttcggtatcaacg 1652
Db 1428 cagttacattcttgcacccaagaagcagccctgaatgcagttgcgtttggactcaacg 1487
Qy 1653 ctgaacaacacacacatactctcttcaggtgataagagacaatgtgatagaccagatag 1712
Db 1488 ccagagaacaacagagaatttcttcgcagggaagaaagactggtcaga---caaatgg 1544
Qy 1713 agaagcaagcagaagatttagcattccctcgtggtgcggtgaaacaagttgaaagcgtcatca 1772
Db 1545 atagcagcgaagaggttatctatttgggtaccatcgaaattggtagataatatattca 1604
Qy 1773 aaaaacagaaggaatctcaactttgtg 1798
Db 1605 acaaccggatgagtcgtatttcag 1630

RESULT 14
AAV72243
ID AAV72243 standard; cDNA; 1924 BP.
XX
AC AAV72243;
XX
DT 24-MAY-1999 (first entry)
XX
DE G. max SBP2 cDNA.
XX
KW SBP1; sucrose binding protein; SBP2; sucrose uptake; transgenic plant;
KW seed; carbohydrate content; soybean; ds.
XX
OS Glycine max.
XX
FH Key Location/Qualifiers
FT 119..1588
FT CDS /*Tag= a
FT /*product= "SBP2"
XX
PN WO9853086-A1.
XX
PD 26-NOV-1998.
XX
PF 21-MAY-1998; 98WO-US10465.
XX
PR 22-MAY-1997; 97US-0047568.
XX
PA (UNIW) UNIV WASHINGTON STATE RES FOUND.
XX
PI Chao WS, Grimes HD;
XX
DR WPI; 1999-070155/06.
DR P-PSDB; AAW90341.
XX
PT New modified plant sucrose binding proteins - used to develop
PT transgenic plants which can have enhanced or decreased sucrose
PT uptake activity in developing seeds
XX
PS Disclosure; Page 40-42; 58pp; English.
XX
CC This sequence encodes a novel sucrose binding protein, SBP2 isolated
CC from Glycine max. The SBP2 protein is used in a method resulting in the
CC production of a modified plant sucrose binding protein (SBP) which has a
CC modified amino acid sequence compared to a corresponding wild-type SBP,
CC and where expression of the modified SBP in a yeast assay system confers

CC enhanced sucrose uptake compared to the corresponding wild-type SBP.
CC The products of the invention can be used for producing transgenic plants
CC which have modified sucrose uptake activity, particularly in developing
CC seeds. Enhanced sucrose uptake activity in developing seeds may be
CC desirable where it is an advantage to increase the carbohydrate content
CC of the seed (e.g. where the seed is the primary plant material harvested,
CC such as soybean). In contrast, decreased sucrose uptake activity in
CC seeds might be desirable where the vegetative material of the plant is
CC harvested. The SBP regulatory regions confer specific or enhanced
CC expression in developing seeds and so may be used to express any
CC transgene in developing seeds.

XX Sequence 1924 BP; 548 A; 432 C; 471 G; 473 T; 0 other;

Query Match 4.9%; Score 98.8; DB 20; Length 1924;
Best Local Similarity 50.2%; Pred. No. 1e-18;
Matches 244; Conservative 0; Mismatches 242; Indels 0; Gaps 0;

Qy 593 tttagacccgcgtacgggaacccaaacggttagatccggtcctgcagaggtttgaccaa 652
Db 437 ttctgcaccagagtcgaacacagaaggtggcagcattcgggtctcagaagttcactgag 496
Qy 653 aggtcaagcaggtttcagaatctccagaatcacgctattgtgcagatcgagcgcaacct 712
Db 497 aaatccaaagcttctcgaagcattgagaatttccgtttggccatcttagaagctagaca 556
Qy 713 aacactctgttcttcccaagcgcgtgatgtgataacatctctgttatccagcaagg 772
Db 557 cacagttcgtgtcccccacgccactttgattccgaggtgtctgttccaacattaagggg 616
Qy 773 caagccaccgtgaccgtagcaaatggcaataacagaaagagctttaaatttgacgagggc 832
Db 617 agagccgtacttgggttggtagggaaagtgaacagaaataccctctagaaacctgga 676
Qy 833 catgcactcagaatcccatccggtttcatttctctacatcttgaccgcctatgacaaccag 892
Db 677 gacatgatacacataccagcagcgcacaccactgtacatcgttaacagagatgagaatgag 736
Qy 893 aacotcagagtagtaaaatctccatgcctgttaacacacccggccagtttgaggatttc 952
Db 737 aagctctcttgcctgcctccatatacctgtctctactctctggaataatttgaggattt 796
Qy 953 ttccgcgcgagcagcagaccatcctactctgcaggtcttcagcaggaataacgttg 1012
Db 797 ttcgggctgagagcagagaccagaatcgtctctcagcattcagcgtgaatgtgctg 856
Qy 1013 gaggccgcttcaatgcgggaattcaatgagatacggagggtgctgttagaagagaaatgca 1072
Db 857 caagctgcgtcccaaccccccaaaaggaagtagaaaggcttttttaatacaacagaacag 916
Qy 1073 ggaagt 1078
Db 917 ggaagt 922

RESULT 15
AAF58252/c
ID AAF58252 standard; DNA; 936 BP.
XX
AC AAF58252;
XX
DT 24-APR-2001 (first entry)
XX
DE Oligonucleotide D1835.
XX
KW Electron-transfer group; ETM; mismatch; genotyping;
KW gene expression; ss.
XX
OS Synthetic.
XX
PN WO200107665-A2.
XX

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 27, 2001, 13:04:45 ; Search time 70.16 Seconds
(without alignments)
218.546 Million cell updates/sec

Title: US-09-715-036-2
1114
Sequence: 1 MAKLTILVALFLLAHAS.....LPRADGDGWFPSVAVDCSG 207

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_1101.*
1: /SIDS2/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1114	100.0	207	22	AA82383
2	824	74.0	156	20	AA40373
3	824	74.0	157	18	AAW24164
4	824	74.0	157	18	AAW24153
5	824	74.0	157	20	AA15245
6	824	74.0	157	22	AAU04707
7	753	67.6	166	20	AA40968
8	683	61.3	167	22	AAU04710
9	681	61.1	166	21	AA333600
10	681	61.1	166	22	AAU05035
11	253	22.7	158	18	AAW23419

12	253	22.7	158	20	AA05723
13	251.5	22.6	155	18	AAW23418
14	239	21.5	158	18	AAW23420
15	162.5	14.6	158	18	AAW23586
16	157.5	14.1	158	18	AAW23588
17	156	14.0	28	18	AAW24194
18	150.5	13.5	155	18	AAW10569
19	150.5	13.5	155	18	AAW23587
20	147.5	13.2	164	10	AAW1892
21	147.5	13.2	164	10	AAW1892
22	142	12.7	184	12	AAW14144
23	142	12.7	184	12	AAW14193
24	142	12.7	184	15	AAW54981
25	142	12.7	184	16	AAW76694
26	142	12.7	184	16	AAW15415
27	142	12.7	184	19	AAW30526
28	142	12.7	184	21	AAW49926
29	140	12.6	140	10	AAW11140
30	138	12.4	140	10	AAW11891
31	137	12.3	28	18	AAW24199
32	137	12.3	184	19	AAW30670
33	135	12.1	189	15	AAW54980
34	129	11.6	171	15	AAW53579
35	128	11.5	110	15	AAW62762
36	128	11.5	111	19	AAW86264
37	128	11.5	149	19	AAW86272
38	127	11.4	149	19	AAW86273
39	125.5	11.3	22	21	AAW27527
40	124	11.1	142	21	AAW84356
41	120.5	10.8	149	19	AAW86274
42	119	10.7	174	22	AAW72901
43	119	10.7	20	21	AAW23115
44	119	10.7	20	21	AAW33564
45	119	10.7	20	22	AAU04777
			20	22	AAU05102

ALIGNMENTS

RESULT	1
AA82383	AA82383 standard; Protein; 207 AA.
XX	XX
AC	AA82383;
XX	XX
DT	23-JUL-2001 (first entry)
XX	XX
DE	Peanut allergen Ara h2 gene product.
XX	XX
KW	Peanut; allergen; Ara h2; transgenic plant; allergy.
XX	XX
OS	Arachis hypogaea.
XX	XX
FH	Key
FT	Peptide
FT	Peptide
FT	Protein
FT	Protein
XX	XX
PN	WO200136621-A2.
XX	XX
PD	25-MAY-2001.
XX	XX
PF	20-NOV-2000; 2000WO-US31657.
XX	XX
PR	19-NOV-1999; 99US-0167255.
XX	XX
PA	(UYAL-) UNIV ALABAMA A & M.
XX	XX
PI	Dodo HW, Arntzen CJ, Konan KN, Viquez OM;
XX	XX
DR	WPI: 2001-355630/37.
DR	N-PSDB; AAF90336, AAF90337.

Soybean Gm2S-1 sul
Soybean albumin 1
Chimeric Soybean a
Mablinin MBLII from
Mablinin MBLIII fr
Peanut allergen Ar
Mablinin II prepro
Mablinin MBLII fro
Amino acid sequenc
Sequence encoded b
Napiin 1-2 gene pro
Napiin protein from
B. campestris napi
Napiin encoded by c
Brassica campestri
Brassica campestri
Sequence encoded b
Part of the sequen
Peanut allergen Ar
Napiin seed storage
Napiin protein from
Synthetic 2S seed
Synthetic 12 kd pr
Wild-type Brazil n
Modified Brazil nu
Modified Brazil nu
Ara h 2 residues 5
Amino acid sequenc
Modified Brazil nu
Flax 2S storage pr
Peanut wild-type A
5/20 native Ara h
Native anaphylacti
Anaphylactic peanu

XX Producing transgenic peanut plants that produce allergen-free seeds,
PT useful in non-allergenic foods, by antisense or sense co-suppression of
PT allergen-encoding genes -
XX
PS Example 1; Fig 2; 72pp; English.
XX
CC The present sequence is that of the peanut allergenic protein (AP)
CC encoded by the Ara h2 gene (see AAF90336). The invention relates
CC to a method for producing a peanut plant having reduced, or
CC undetectable, AP content in its seed. A peanut plant cell is
CC transformed with a DNA construct containing an antisense AP gene
CC and/or sense AP gene, or their fragments, regenerated to plants,
CC and fertile transgenic plants that produce seeds with reduced AP
CC content are identified. The AP sense or antisense gene may
CC comprise at least a portion of the Ara h2 gene sequence. The seeds
CC are useful for preparation of allergen-free foods. Recombinant AP
CC may be produced and used to produce antibodies useful for detecting
CC AP in foods, and for treatment or prevention of peanut allergy.
XX
SQ Sequence 207 AA;

Query Match 100.0%; Score 1114; DB 22; Length 207;
Best Local Similarity 100.0%; Pred. No. 4.8e-110;
Matches 207; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MAKITILVALFLLAHASARQWELQGRRCQSLERANLRPCQHLMKIQRDEDSY 60
DB 1 makitilvalflllaahasargqwelqgdrccqlerlanlrpcqhlmqkigrded 60
QY 61 ERDPYSPQDPYSPYDRRGAGSSQHQERCCNELNENFENNCRCMCEALQOIMENQSDRL 120
DB 61 erdpyspqdypspypdyrrgagssqhqerccnelnefenncrcmcealqqimengsdr 120
QY 121 QGRQEQFKRELRLNPQCGLRAPQRCDLQVSGRRPRIPPLTGSRRRHQSPVGNR 180
DB 121 qgrqeqqfkelrnlpqcgrlrapqrcdlqvesgrrprippiltgsrrrrhqsypgnr 180
QY 181 RYSACMLLPRAADGWFPSVAVDCSG 207
DB 181 rysacmlpraadgdgwfpsvavdcsg 207

RESULT 2
AA40973 2
ID AA40973 standard; protein; 156 AA.

XX AC AA40973
XX DT 06-DEC-1999 (first entry)
XX PE Peanut h 2 protein fragment.
XX KW Peanut; allergen; Ara h 1; IgE; immunoglobulin E; epitope; Ara h 3;
XX KW allergic reaction; Ara h 2.
XX OS Arachis hypogaea.
XX PN WO9945961-A1.
XX PD 16-SEP-1999.
XX PF 12-MAR-1999; 99WO-US05494.
XX PR 12-MAR-1998; 98US-0077763.
XX PR 11-MAR-1999; 99US-0077763.
XX PA (UYAR-) UNIV ARKANSAS.
XX PI Burks W, Helm RM, Cockrell G, Bannon GA, Stanley JS, Shin DS;
XX PI Sampson H, Compadre CM, Huang SK, Maleki SJ, Kopper RA;

DR WPI; 1999-551218/46.
XX Tertiary structure of peanut allergen Ara h 1 for protection of a host
PT animal from allergic reaction -
XX
PS Disclosure; Page 104; 193pp; English.
XX
CC The invention provides a tertiary structure for the peanut allergen
CC Ara h 1. The Ara h 1 allergen is found to contain 23 linear IgE-binding
CC epitopes. The invention also provides an isolated recombinant peanut
CC allergen designated Ara h 3 and a nucleotide molecule encoding the peanut
CC allergen Ara h 3. Molecules of the invention are used to protect a host
CC animal from allergic reaction, particularly using a modified allergen
CC which is less reactive with IgE. The invention may also be used to
CC ensure that the allergen is not introduced into genetically modified
CC food. The present sequence represents an Ara h 2 protein fragment.
XX
SQ Sequence 156 AA;

Query Match 74.0%; Score 824; DB 20; Length 156;
Best Local Similarity 100.0%; Pred. No. 1.8e-79;
Matches 154; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 4 LITILVALFLLAHASARQWELQGRRCQSLERANLRPCQHLMKIQRDEDSYERD 63
DB 1 litilvalflllaahasargqwelqgdrccqlerlanlrpcqhlmqkigrdedsyerd 60
QY 64 PYSPQDPYSPYDRRGAGSSQHQERCCNELNENFENNCRCMCEALQOIMENQSDRLQGR 123
DB 61 pypsqdypspypdyrrgagssqhqerccnelnefenncrcmcealqqimengsdr 120
QY 124 QEQQFKRELRLNPQCGLRAPQRCDLQVSGRR 157
DB 121 qeqqfkelrnlpqcgrlrapqrcdlqvesgrr 154

RESULT 3
AAW24164
ID AAW24164 standard; Protein; 157 AA.

XX AC AAW24164;
XX DT 29-DEC-1997 (first entry)
XX PE Peanut allergen Ara hII.
XX KW Peanut; seed storage protein; allergen; allergy; hypersensitivity;
XX KW vaccine; anaphylactic shock; immunotherapy; therapy;
XX KW monoclonal antibody; ELISA; analysis; Ara hII.
XX OS Arachis hypogaea strain Florunner.
XX PN WO9724139-A1.
XX PD 10-JUL-1997.
XX PF 23-SEP-1996; 96WO-US15222.
XX PR 04-MAR-1996; 96US-0610424.
XX PR 29-DEC-1995; 95US-0009455.
XX PA (UYAR-) UNIV ARKANSAS.
XX PI Bannon GA, Burks AW, Cockrell G, Helm RM, Stanley JS;
XX WPI; 1997-363453/33.
XX DR N-PSDB; AAT76614.
XX PT Peanut allergens Ara hI and Ara hII - used for vaccination and in
XX PT two-site monoclonal antibody based ELISA
XX PS Claim 31; Page 219; 354pp; English.

XX This polypeptide comprises major peanut allergen Ara hII.
 CC Its sequence was deduced from cDNA clone P38 (AA76614), isolated
 CC from peanut seed cDNA using a primer (see AAT76617) based on an
 CC isolated Ara hI peptide (see AAW24151). The sequence shows
 CC significant homology with the conglutin family of seed storage
 CC proteins of other legumes. The allergen is recognised by serum
 CC IgE from a large proportion of individuals with peanut
 CC hypersensitivity. Ara hII and Ara hI (see AAW24149-50) can be used to
 CC raise monoclonal antibodies which are used in a specific two-site
 CC MAb ELISA for the detection of Ara hI or Ara hII (claimed). IgE-
 CC binding Ara hII antigen epitopes (see AAW24188-93) may be used in
 CC vaccines to protect against allergic reactions to peanut allergens,
 CC e.g. anaphylactic shock.
 XX
 SQ Sequence 157 AA;

Query Match 74.0%; Score 824; DB 18; Length 157;
 Best Local Similarity 100.0%; Pred. No. 1.8e-79;
 Matches 154; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 4 LTIILVALFLLAAHASARQWELQDRRCQSLERANLRPCQHLMQKIQRDEDSYERD 63
 Db 1 LTIILVALFLLAAHASARQWELQDRRCQSLERANLRPCQHLMQKIQRDEDSYERD 60
 QY 64 PYSPQDPYSPYDRRGAGSSQHQERCCNELNEFENNORCMCEALQOIMENQSDRLQGR 123
 Db 61 PYSPQDPYSPYDRRGAGSSQHQERCCNELNEFENNORCMCEALQOIMENQSDRLQGR 120
 QY 124 QOEOQFKRELNLPOQCGLRAPQRCDLVDSGGR 157
 Db 121 QOEOQFKRELNLPOQCGLRAPQRCDLVDSGGR 154

RESULT 4
 AAW24153
 ID AAW24153 standard; Protein; 157 AA.
 XX
 AC AAW24153.
 DT 29-DEC-1997 (first entry)
 XX Peanut allergen Ara hII.
 DE Peanut; seed storage protein; allergen; allergy; hypersensitivity;
 KW vaccine; anaphylactic shock; immunotherapy; therapy;
 KW monoclonal antibody; ELISA; analysis; Ara hII.
 XX
 OS Arachis hypogaea strain Florunner.
 XX WO9724139-A1.
 PN 10-JUL-1997.
 PD 23-SEP-1996; 96WO-US15222.
 XX 04-MAR-1996; 96US-0610424.
 PR 29-DEC-1995; 95US-0009455.
 XX (UYAR-) UNIV ARKANSAS.
 PA Bannon GA, Burks AW, Cockrell G, Helm RM, Stanley JS;
 PI WPI: 1997-363453/33.
 DR N-PSDB; AAT76615.
 XX Peanut allergens Ara hI and Ara hII - used for vaccination and in
 PT two-site monoclonal antibody based-ELISA
 XX Claim 31; Page 198; 354pp; English.
 PS This polypeptide comprises major peanut allergen Ara hII.
 XX

CC Its sequence was deduced from a cDNA clone (AAT76615) isolated
 CC from peanut seed cDNA using a primer (see AAT76617) based on an
 CC isolated Ara hI peptide (see AAW24151). The sequence shows
 CC significant homology with the conglutin family of seed storage
 CC proteins of other legumes. The allergen is recognised by serum
 CC IgE from a large proportion of individuals with peanut
 CC hypersensitivity. Ara hII and Ara hI (see AAW24149-50) can be used to
 CC raise monoclonal antibodies which are used in a specific two-site
 CC MAb ELISA for the detection of Ara hI or Ara hII (claimed). IgE-
 CC binding Ara hII antigen epitopes (see AAW24188-93) may be used in
 CC vaccines to protect against allergic reactions to peanut allergens,
 CC e.g. anaphylactic shock.
 XX
 SQ Sequence 157 AA;

Query Match 74.0%; Score 824; DB 18; Length 157;
 Best Local Similarity 100.0%; Pred. No. 1.8e-79;
 Matches 154; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 4 LTIILVALFLLAAHASARQWELQDRRCQSLERANLRPCQHLMQKIQRDEDSYERD 63
 Db 1 LTIILVALFLLAAHASARQWELQDRRCQSLERANLRPCQHLMQKIQRDEDSYERD 60
 QY 64 PYSPQDPYSPYDRRGAGSSQHQERCCNELNEFENNORCMCEALQOIMENQSDRLQGR 123
 Db 61 PYSPQDPYSPYDRRGAGSSQHQERCCNELNEFENNORCMCEALQOIMENQSDRLQGR 120
 QY 124 QOEOQFKRELNLPOQCGLRAPQRCDLVDSGGR 157
 Db 121 QOEOQFKRELNLPOQCGLRAPQRCDLVDSGGR 154

RESULT 5
 AAY15245
 ID AAY15245 standard; Protein; 157 AA.
 XX
 AC AAY15245.
 DT 09-NOV-1999 (first entry)
 XX Peanut allergen, Ara h 2, amino acid sequence.
 DE Allergy; immune response; transgenic; allergen; epitope;
 KW immunoglobulin E; Ig E; binding site; peanut.
 XX
 OS Arachis hypogaea.
 XX WO9938978-A1.
 PN 05-AUG-1999.
 PD 29-JAN-1999; 99WO-US02031.
 PF 27-AUG-1998; 98US-0141220.
 XX 31-JAN-1998; 98US-0073283.
 PR 13-FEB-1998; 98US-0074590.
 PR 13-FEB-1998; 98US-0074624.
 PR 13-FEB-1998; 98US-0074633.
 XX (SOSI/) SOSIN H.
 PA (UYAR-) UNIV ARKANSAS.
 PI (UYNY) UNIV NEW YORK MT SINAI SCHOOL MEDICINE.
 DR Bannon GA, Burks AW, Sampson HA, Sosin H;
 DR WPI: 1999-479189/40.
 DR N-PSDB; AAZ06383.
 XX Modified allergen with reduced IgE binding, useful for treating e.g.
 PT allergies
 PS Disclosure; Page 38; 46pp; English.
 XX

XX This is the amino acid sequence of the Ara h 2 protein from Arachis
CC hypogaea. The Ara h 2 protein has 10 IgE (immunoglobulin E) binding
CC epitopes, three of which are immunodominant (AAV15272, AAV15275, and
CC AAV15276).
CC By modifying the IgE binding sites the ability of the allergen to
CC provoke an immune response is downregulated. The epitopes of the IgE
CC binding sites can therefore be modified in genetically engineered plants
CC and animals to elicit less of an allergic response.
XX
SQ Sequence 157 AA;

Query Match 74.0%; Score 824; DB 20; Length 157;
Best Local Similarity 100.0%; Pred. No. 1.8e-79;
Matches 154; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 LTIIVALLFLAAHASARQWELQGRRCOSQLERANLRPCQHLMQKIQRDEDSYERD 63
DB 1 LTIIVALLFLAAHASARQWELQGRRCOSQLERANLRPCQHLMQKIQRDEDSYERD 60

QY 64 PYPSPQDPSPYDRRGSSQHQRCCNELNFENNRCMCCEALQOIMENOSDRLOQR 123
DB 61 PYPSPQDPSPYDRRGSSQHQRCCNELNFENNRCMCCEALQOIMENOSDRLOQR 120

QY 124 QEOQFKRELNLPOQCGLRAPQRCDLVDSGGR 157
DB 121 QEQQFKRELNLPOQCGLRAPQRCDLVDSGGR 154

RESULT 6
AAU04707
ID AAU04707 standard; Protein; 157 AA.
XX
AC AAU04707;
XX
XX 23-OCT-2001 (first entry)
XX
DE Anaphylactic antigen Ara h 2.
XX
XX Ara h 2; anaphylactic antigen; immunoglobulin E; IgE; immunogenic;
KW allergy; mast cell; basophil; mouse.
XX
OS Mus sp.
XX
XX WO200140264-A2.
XX
XX 07-JUN-2001.
XX
XX 06-DEC-2000; 2000WO-US33124.
XX
XX 06-DEC-1999; 99US-0455294.
XX
XX 23-JUN-2000; 2000US-0213765.
XX
XX 27-SEP-2000; 2000US-0235797.
XX
XX (PANA-) PANACEA PHARM LLC.
XX
XX (UYAR-) UNIV ARKANSAS.
XX
XX (MOUN) MOUNT SINAI SCHOOL MEDICINE NEW YORK.
XX
XX Bannon GA, Burks WA, Caplan MJ, Sampson H, Sosin H;
XX
XX WPI; 2001-381378/40.
XX
XX Antigenic fragments useful for reducing anaphylactic risk and reducing
PT the severity and/or number of allergic symptoms in individuals
PT sensitive to antigens, have reduced ability to bind immunoglobulin E -
XX
XX Claim 7; Fig 10; 100pp; English.
XX
XX The sequence represents the amino acid sequence of anaphylactic antigen
CC Ara h 2. Ara h 2 is an anaphylactic antigen (A), which was used to design
CC antigenic peptides having a reduced ability to bind immunoglobulin E
CC (IgE) as compared with the intact (A), or having a sequence substantially

CC identical to a portion of sequence of an antigen that includes at least
CC one IgE binding site, where at least one IgE binding site of the peptide
CC is altered. The antigenic peptides are used in a composition which is
CC useful for reducing risk or severity of allergic reaction to an antigen.
CC This is done by identifying an individual at risk of allergic reaction to
CC an antigen by identifying prior display of allergic symptoms when exposed
CC to the antigen, or a familial relationship with an individual who
CC previously displayed allergic symptoms when exposed to the antigen.
CC Following this an antigen-specific IgE present on one or more mast cells
CC or basophils in the individual's serum is identified. The individual is
CC then contacted with a peptide corresponding to a portion of the
CC antigen, which is selected, formulated, and delivered so that binding of
CC the peptide to antigen-specific IgE is reduced as compared with IgE
CC binding of intact antigen. The composition is also useful for treating
CC and preventing allergic reactions.
XX
SQ Sequence 157 AA;

Query Match 74.0%; Score 824; DB 22; Length 157;
Best Local Similarity 100.0%; Pred. No. 1.8e-79;
Matches 154; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 LTIIVALLFLAAHASARQWELQGRRCOSQLERANLRPCQHLMQKIQRDEDSYERD 63
DB 1 LTIIVALLFLAAHASARQWELQGRRCOSQLERANLRPCQHLMQKIQRDEDSYERD 60

QY 64 PYPSPQDPSPYDRRGSSQHQRCCNELNFENNRCMCCEALQOIMENOSDRLOQR 123
DB 61 PYPSPQDPSPYDRRGSSQHQRCCNELNFENNRCMCCEALQOIMENOSDRLOQR 120

QY 124 QEOQFKRELNLPOQCGLRAPQRCDLVDSGGR 157
DB 121 QEQQFKRELNLPOQCGLRAPQRCDLVDSGGR 154

RESULT 7
AAV40968
ID AAV40968 standard; protein; 166 AA.
XX
AC AAV40968;
XX
XX 06-DEC-1999 (first entry)
XX
XX Recombinant Ara h 2 protein sequence.
XX
XX Peanut; allergen; Ara h 1; IgE; immunoglobulin E; epitope; Ara h 3;
KW allergic reaction; Ara h 2.
XX
XX Arachis hypogaea.
XX
XX WO9945961-A1.
XX
XX 16-SEP-1999.
XX
XX 12-MAR-1999; 99WO-US05494.
XX
XX 12-MAR-1998; 98US-0077763.
XX
XX 11-MAR-1999; 99US-0077763.
XX
XX (UYAR-) UNIV ARKANSAS.
XX
XX Burks W, Helm RM, Cockrell G, Bannon GA, Stanley JS, Shin DS;
PI Sampson H, Compadre CM, Huang SK, Maleki SJ, Kopper RA;
XX
XX WPI; 1999-551218/46.
XX
XX Tertiary structure of peanut allergen Ara h 1 for protection of a host
PT animal from allergic reaction -
XX
XX Disclosure; Page 81; 193pp; English.
XX
XX The invention provides a tertiary structure for the peanut allergen
CC

CC Ara h 1. The Ara h 1 allergen is found to contain 23 linear IgE-binding
CC epitopes. The invention also provides an isolated recombinant peanut
CC allergen designated Ara h 3 and a nucleotide molecule encoding the peanut
CC allergen Ara h 3. Molecules of the invention are used to protect a host
CC animal from allergic reaction, particularly using a modified allergen
CC which is less reactive with IgE. The invention may also be used to
CC ensure that the allergen is not introduced into genetically modified
CC food. The present sequence represents a recombinant Ara h 2 protein.
XX
SQ Sequence 166 AA;

Query Match 67.6%; Score 753; DB 20; Length 166;
Best Local Similarity 100.0%; Pred. No. 6.5e-72;
Matches 138; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 20 SARQWELQGRRCOSOLERANLRPCQHLMOKIQRDSEYERDPYSPSPSPYDR 79
DQ 15 SARQWELQGRRCOSOLERANLRPCQHLMOKIQRDSEYERDPYSPSPSPYDR 74
QY 80 RGAGSSQHQERCCNELEFENNORCMCEALQIQIMENQSDRLQGRQEQQFKRELNLPPQ 139
DQ 75 RGAGSSQHQERCCNELEFENNORCMCEALQIQIMENQSDRLQGRQEQQFKRELNLPPQ 134
QY 140 CGLRAPQRCDDLVSQGR 157
DQ 135 CGLRAPQRCDDLVSQGR 152

RESULT 8
AAU04710
ID AAU04710 standard; Protein; 167 AA.
XX
AC AAU04710;
XX
DT 23-OCT-2001 (first entry)
XX
DE Modified anaphylactic antigen Ara h 2.
XX
DE Ara h 2; anaphylactic antigen; immunoglobulin E; IgE; immunogenic;
KW allergy; mast cell; basophil; mouse.
XX
OS Mus sp.
OS Synthetic.
XX
PN WO200140264-A2.
XX
PD 07-JUN-2001.
XX
PF 06-DEC-2000; 2000WO-US33124.
XX
PR 06-DEC-1999; 99US-0455294.
PR 23-JUN-2000; 2000US-0213765.
PR 27-SEP-2000; 2000US-0235797.
XX
PA (PANA-) PANACEA PHARM LLC.
PA (UYAR-) UNIV ARKANSAS.
PA (MOUN) MOUNT SINAI SCHOOL MEDICINE NEW YORK.
XX
PI Bannon CA, Burks WA, Caplan MJ, Sampson H, Sosin H;
XX
XX WPI; 2001-381378/40.
XX
PT Antigenic fragments useful for reducing anaphylactic risk and reducing
PT the severity and/or number of allergic symptoms in individuals
PT sensitive to antigens, have reduced ability to bind Immunoglobulin E -
XX
PS Disclosure; Fig 12; 100pp; English.
XX
XX The sequence represents the amino acid sequence of modified anaphylactic
CC antigen Ara h 2, which has been altered to disrupt immunoglobulin E (IgE)
CC binding sites. Ara h 2 is an anaphylactic antigen (A), which was used to
CC design antigenic peptides having a reduced ability to bind IgE as

CC compared with the intact (A), or having a sequence substantially
CC identical to a portion of sequence of an antigen that includes at least
CC one IgE binding site, where at least one IgE binding site of the peptide
CC is altered. The antigenic peptides are used in a composition which is
CC useful for reducing risk or severity of allergic reaction to an antigen.
CC This is done by identifying an individual at risk of allergic reaction to
CC an antigen by identifying prior display of allergic symptoms when exposed
CC to the antigen, or a familial relationship with an individual who
CC previously displayed allergic symptoms when exposed to the antigen.
CC Following this an antigen-specific IgE present on one or more mast cells
CC or basophils in the individual's serum is identified. The individual is
CC then contacted with a peptide corresponding to a portion of the
CC antigen, which is selected, formulated, and delivered so that binding of
CC the peptide to antigen-specific IgE is reduced as compared with IgE
CC binding of intact antigen. The composition is also useful for treating
CC and preventing allergic reactions.
XX
SQ Sequence 167 AA;

Query Match 61.3%; Score 683; DB 22; Length 167;
Best Local Similarity 91.5%; Pred. No. 1.7e-64;
Matches 129; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 20 SARQWELQGRRCOSOLERANLRPCQHLMOKIQRDSEYERDPYSPSPSPYDR 79
DQ 15 SARQWELQGRRCOSOLERANLRPCQHLMOKIQRDSEYERDPYSPSPSPYDR 74
QY 80 RGAGSSQHQERCCNELEFENNORCMCEALQIQIMENQSDRLQGRQEQQFKRELNLPPQ 139
DQ 75 RGAGSSQHQERCCNELEFENNORCMCEALQIQIMENQSDRLQGRQEQQFKRELNLPPQ 134
QY 140 CGLRAPQRCDDLVSQGR 160
DQ 135 CGLRAPQRCDDLVSQGR 155

RESULT 9
AAB33600
ID AAB33600 standard; Protein; 166 AA.
XX
AC AAB33600;
XX
DT 22-JAN-2001 (first entry)
XX
DE Modified Ara h 2 amino acid sequence.
XX
KW Allergy; allergic reaction; allergen; anaphylactic antigen; peanut;
KW Ara h 1; Ara h 2; Ara h 3; epitope; binding; immunoglobulin E; IgE;
KW food antigen; sensitising; immune response; anti-allergic.
XX
OS Arachis hypogaea.
XX
PN WO200051647-A2.
XX
PD 08-SEP-2000.
XX
PF 03-MAR-2000; 2000WO-US05655.
XX
PR 03-MAR-1999; 99US-0122960.
PR 06-DEC-1999; 99US-0455294.
XX
PA (MOUN) MOUNT SINAI SCHOOL MEDICINE.
XX
PI Sampson HA;
XX
DR WPI; 2000-611341/58.
XX
PT Non-human animal sensitized to an antigen, useful as an animal model
PT for studying allergic reactions to allergens, such as those in food and
PT in the environment -
XX
PS Example 6; Fig 17B; 124pp; English.

XX The present invention describes an animal model which can be used for
CC studying allergic reactions to allergens. The animal is sensitised to
CC a selected antigen by administering the antigen itself or a nucleic
CC acid encoding the antigen, where preferably the antigen is an
CC anaphylactic antigen. The sensitised animal can then be used to screen
CC for compounds which may help to prevent, ameliorate, or cure allergic
CC conditions in humans. The animal model can be used for studying allergic
CC reactions to allergens, such as those in food (peanuts, fruit, berry,
CC nut, bean, milk, dairy products), or in the environment (weed pollen,
CC grass pollen, tree pollen, mite, animal, animal dander, fungal, and
CC insect antigens). AAB33478 to AAB33601 represent sequences which are
CC used in examples from the present invention to specifically examine the
CC peanut allergy, and the peanut anaphylactic antigens Ara h 1 to 3.
XX
SQ Sequence 166 AA;

Query Match 61.1%; Score 681; DB 21; Length 166;
Best Local Similarity 92.8%; Pred. No. 2.8e-64;
Matches 128; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 20 SARQWELQGRDRCQSLERANLRPCQHLQMKIQRDSDSYERDPYSPQDPYSPYDR 79
DB 15 sarqaelqgdrccqqlaranlraceahlmqkiqadedsyerapyspsqapyspydr 74
QY 80 RGAGSSQHQRCCNELNEFENNRCMCEALQQTIMENOSDRLOGRQEQQFKRELRLNPQQ 139
DB 75 rgagssqhqrccnelnefennqrcmcealqgimengsdrlogaqeqqfcrearnlpq 134
QY 140 CGLRAPQRCDLVDSEGG 157
DB 135 cglrapqrcdadvessgr 152

RESULT 10
AAU05035
ID AAU05035 standard; protein; 166 AA.
AC AAU05035;
XX
XX 24-OCT-2001 (first entry)
XX Modified anaphylactic peanut antigen Ara h 2.
DE Anaphylactic antigen; Ara h 2; peanut; desensitisation; antigen;
XX allergy; immunotherapy; rush immunotherapy; anaphylaxis; asthma;
KW food allergen; milk allergen; shellfish allergen; latex; drug;
KW environmental allergen; grass pollen; ovalbumin; insect venom;
KW peanut oil; immunoglobulin E; IgE.
XX
XX Arachis sp.
OS Synthetic.
OS
XX WO200139799-A2.
PN
XX 07-JUN-2001.
PD
XX 06-DEC-2000; 2000WO-US33125.
PF
XX 06-DEC-1999; 99US-0455294.
PR 23-JUN-2000; 2000US-0213765.
PR 27-SEP-2000; 2000US-0235797.
XX
XX (PANA-) PANACEA PHARM LLC.
PA
XX Caplan M;
PI
XX WPI; 2001-408258/43.
XX
XX Preventing allergic response against antigens, e.g. food and
PT environmental allergens such as peanut allergen or grass pollen, latex
PT or drug, comprises administering agent, e.g. a peptide, that blocks

PT antigen binding sites on offending IgE -
XX Disclosure; Fig 2; 76pp; English.
PS
XX The sequence represents the amino acid sequence of modified anaphylactic
CC peanut antigen Ara h 2. The protein is used for active or passive
CC desensitisation of an individual to an antigen; for alleviating or
CC preventing allergic reactions and for decreasing the risk of allergic
CC reactions during immunotherapy or rush immunotherapy, anaphylaxis
CC and asthma. The antigen may be a food allergen (e.g. peanut or milk
CC allergen), shellfish allergen, environmental allergen (e.g. grass pollen
CC or tree pollen), latex, drug, pollen, ovalbumin, an insect venom antigen
CC or predominantly linear epitopes. The protein is useful for protecting an
CC individual against subsequent inadvertent or intentional exposure to
CC antigen, e.g. receiving blocking agent before eating a chocolate bar
CC which may inadvertently contain peanut components or before eating foods
CC prepared using peanut oil. Administration of the blocking agents does not
CC result in cross-linking of anti-antigenic immunoglobulin E (IgE). After
CC exposure to the agent, the individual's antigen sensitivity is at least
CC temporarily reduced. Only those IgE molecules that bind the offending
CC antigen and contribute to the risk of an allergic response are blocked.
XX
SQ Sequence 166 AA;

Query Match 61.1%; Score 681; DB 22; Length 166;
Best Local Similarity 92.8%; Pred. No. 2.8e-64;
Matches 128; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 20 SARQWELQGRDRCQSLERANLRPCQHLQMKIQRDSDSYERDPYSPQDPYSPYDR 79
DB 15 sarqaelqgdrccqqlaranlraceahlmqkiqadedsyerapyspsqapyspydr 74
QY 80 RGAGSSQHQRCCNELNEFENNRCMCEALQQTIMENOSDRLOGRQEQQFKRELRLNPQQ 139
DB 75 rgagssqhqrccnelnefennqrcmcealqgimengsdrlogaqeqqfcrearnlpq 134
QY 140 CGLRAPQRCDLVDSEGG 157
DB 135 cglrapqrcdadvessgr 152

RESULT 11
AAW23419
ID AAW23419 standard; protein; 158 AA.
AC AAW23419;
XX
XX 08-APR-1998 (first entry)
DT
XX Soybean albumin 3 protein.
DE
XX Soybean albumin 3; seed storage protein; SSP; transgenic seed;
KW transgenic soybean plant; animal feed production.
KW
XX Glycine max.
OS
XX
XX Key Location/Qualifiers
FH Peptide 1..21
FT /note= "prepropeptide"
FT Protein 22..158
FT /note= "mature protein"
FT Misc-difference 22..81
FT /note= "small chain sequence"
FT Misc-difference 81..158
FT /note= "large chain sequence"
XX
XX WO9735023-A2.
PN
XX 25-SEP-1997.
PD
XX 19-MAR-1997; 97WO-US04409.
PF
XX

PR 20-MAR-1996; 96US-0618911.
 XX (PION-) PIONEER HI-BRED INT INC.
 PA Coughlan S, Hastings C, Hu D, Jung R;
 PI WPI; 1997-480228/44.
 XX DR N-PSDB; AAT86282.
 PT Increasing the nutritional quality of soybean - by genetically
 PT increasing the amount of lysine, methionine and/or cysteine in
 PT albumin proteins
 XX Claim 25; Page 46-47; 62pp; English.
 XX This sequence represents the soybean albumin 3 protein. DNA encoding this
 CC sequence represents a DNA molecule of the invention. The DNA molecules of
 CC the invention comprise a preselected DNA segment encoding a seed storage
 CC protein (SSP). (1) is used to produce transgenic seeds and plants,
 CC especially soybean plants that have increased levels of lysine and also
 CC increased levels of methionine and/or cysteine in albumin protein type 1
 CC and/or 3. Increasing the levels of these amino acids increases the
 CC nutritional value of soybean produced. This is especially useful for
 CC producing animal feeds. The amount of lysine in the seed is increased by
 CC 5-10 %, the amount of methionine and cysteine is increased by 10-30%
 XX Sequence 158 AA;
 SQ
 Query Match 22.7%; Score 253; DB 18; Length 158;
 Best Local Similarity 36.7%; Pred. No. 5.9e-19;
 Matches 58; Conservative 31; Mismatches 57; Indels 12; Gaps 6;
 QY 1 MAKLTILVALALFLAAHASARQWELQGRRCOSQLERANLRPCEOHLMOKIQ--RDDE 58
 Db 1 mtkftillilslfci-ahcscaskwqhqqd-scrkqlggvnltpcekhimekiqgrgdd 58
 QY 59 SYERDPYSPDYPSPSPYDRRGAGSSQ-----HQERCCNELNEFNQRCMCALQOI 112
 Db 59 ddddddnhilrtmgrinyirnegkdedeeeghmqkctemsel-rspkqcqkalki 117
 QY 113 MENQSDRLQGRQEQQFKRELNLPPQCGLRAPQRCDL 150
 Db 118 menqseele-ekqkkmekelinlatmcrfgpmiqcdl 154
 RESULT 12
 AAY05723
 XX AAY05723 standard; Protein; 158 AA.
 XX 19-JUL-1999 (first entry)
 XX Soybean Gm2S-1 sulphur-rich cotyledon-specific protein.
 DE Gm2S-1; lunasin; alisin; soybean; albumin; cotyledon; mitosis;
 KW mitotic disruptor; cancer; restenosis; therapy.
 XX Glycine max.
 XX Key Location/Qualifiers
 FH Peptide 1..21
 FT /note= "signal peptide"
 FT Protein 22..64
 FT /note= "lunasin"
 FT Protein 65..81
 FT /note= "linker peptide"
 FT Protein 82..158
 FT /note= "alisin"
 PN W09915642-A1.
 XX

PD 01-APR-1999.
 XX 25-SEP-1998; 98WO-US201116.
 PF 25-SEP-1997; 97US-0938675.
 XX (REGC) UNIV CALIFORNIA.
 PA De Lumen BO, Galvez AF;
 XX WPI; 1999-254704/21.
 XX DR N-PSDB; AAX25346.
 XX Soybean cotyledon-specific 2S albumin cDNA as a mitotic disruptor
 PT Claim 5; Page 37; 39pp; English.
 XX This sequence represents soybean cotyledon-specific sulphur-rich
 CC protein Gm2S-1 precursor, as encoded by a cDNA clone (see AAX25346)
 CC isolated from a mid-maturation seed cDNA library of soybean cv.
 CC Hodgson 78. The protein precursor is composed of a 21-amino acid
 CC residue signal peptide and a pro-protein that is post-translationally
 CC processed to yield a 43 amino acid small subunit (lunasin) with a
 CC unique C-terminal end containing the RGD cell adhesion motif
 CC followed by 8 Asp residues, a 17-amino acid linker peptide, and a
 CC 77 amino acid (8 kDa) large subunit (alisin) that is rich in sulphur
 CC (7.8% methionine) in lysine (13.0%) and cysteine (7.8%). The
 CC invention provides methods and compositions for selectively
 CC modulating mitotic function in a target cell, including mammalian,
 CC plant and bacterial cells, in vitro or in situ. The methods involve
 CC contacting the cell with a modulator of mitotic function comprising
 CC a Gm2S-1 peptide, particularly a lunasin and/or alisin peptide. The
 CC peptide may be introduced by transfecting the cell with a nucleic
 CC acid encoding the peptide, or a fragment which modulates expression
 CC of a resident Gm2S-1 peptide-encoding gene or transcript. Mentioned
 CC in the disclosure are the control of undesirable cell growth in human
 CC neoplasia e.g. cancer, restenosis, etc., prevention of division of
 CC pathogenic bacteria and the regulation of plant seed development.
 XX Sequence 158 AA;
 SQ
 Query Match 22.7%; Score 253; DB 20; Length 158;
 Best Local Similarity 36.7%; Pred. No. 5.9e-19;
 Matches 58; Conservative 31; Mismatches 57; Indels 12; Gaps 6;
 QY 1 MAKLTILVALALFLAAHASARQWELQGRRCOSQLERANLRPCEOHLMOKIQ--RDDE 58
 Db 1 mtkftillilslfci-ahcscaskwqhqqd-scrkqlggvnltpcekhimekiqgrgdd 58
 QY 59 SYERDPYSPDYPSPSPYDRRGAGSSQ-----HQERCCNELNEFNQRCMCALQOI 112
 Db 59 ddddddnhilrtmgrinyirnegkdedeeeghmqkctemsel-rspkqcqkalki 117
 QY 113 MENQSDRLQGRQEQQFKRELNLPPQCGLRAPQRCDL 150
 Db 118 menqseele-ekqkkmekelinlatmcrfgpmiqcdl 154
 RESULT 13
 AAW23418
 ID AAW23418 standard; Protein; 155 AA.
 XX AAW23418;
 AC AAW23418;
 XX 08-APR-1998 (first entry)
 XX Soybean albumin 1 protein.
 DE Soybean albumin 1 protein.
 KW Soybean albumin 1; seed storage protein; SSP; transgenic seed;
 KW transgenic soybean plant; animal feed production.
 XX Glycine max.
 OS

XX	Key	Location/Qualifiers
FH	Peptide	1..20
FT		/note= "signal peptide"
FT	Protein	21..155
FT		/note= "mature protein"
FT	Misc-difference	22..75
FT		/note= "small chain sequence"
FT	Misc-difference	76..155
FT		/note= "large chain sequence"

WO9735023-A2.
25-SEP-1997.
19-MAR-1997; 97WO-US04409.
20-MAR-1996; 96US-0618911.
(PION-)\ PIONEER HI-BRED INT. INC.

Coughlan S, Hastings C, Hu D, Jung R;
WPI; 1997-480228/44.
N-PSDB: AAT86281.

AA Increasing the nutritional quality of soybean - by genetically
PT increasing the amount of lysine, methionine and/or cysteine in
PT albumin proteins
PT

Claim 25: page 44-45: 62pp: English.

This sequence represents the soybean albumin 1 protein. DNA encoding this sequence represents a DNA molecule of the invention. The DNA molecule of the invention comprises a preselected DNA segment encoding a seed storage protein (SSP). (1) is used to produce transgenic seeds and plants, especially soybean plants that have increased levels of lysine and also increased levels of methionine and/or cysteine in albumin protein type 1 and/or 3. Increasing the levels of these amino acids increases the nutritional value of soybean produced. This is especially useful for producing animal feeds. The amount of lysine in the seed is increased by 5-10 % the amount of methionine and cysteine is increased by 10-30 %

Sequence 155 AA;

Query Match	22.68;	Score 251.5;	DB 18;	Length 155;
Best Local Similarity	38.28;	Pred. NO. 8.3e-19;		
Matches	60;	Conservative	30;	Mismatches 54;
				Indels 13;
				Gaps 7;

Qy	1	MAKLIIIVAlAFLILAHASAROOELOGDRCSQLEFRANLRCEQHLHMKQIO--RDE	57
Db	1	mckllilallfll--antccaskqhghqscreqkqinlncpe-himekqaggrg	57
Qy	58	DSYERDPSPSPDPSPYDRRAGSSQ-----HQBRCSNELNEPENNQRMCZALQOIM	113
Db	58	dgsdedhllrtmp-grfnylrkkgdeeeehgmhgcscemselk-sp1cqckalqkim	115
Qy	114	ENQSRLOGRQEOQFKRELNLTPQOCGLRAPQRCDL	150
Db	116	dnasee1eq-ekkamere1mltalrtqlpmqcdl	151

RESULT 14
AAW23420
ID AAW23420 standard: protein: 158 AA.

AA	AAW23420;
AC	
XX	
XX	08-APR-1998 (first entry)
DT	
XX	
XX	Chimeric Soybean albumin 1/3 protein.
DE	
XX	

Soybean albumin 1; seed storage protein; SSP; transgenic seed; transgenic soybean plant; animal feed production; chimeric protein.

OS Synthetic.
OS Glycine max.

PN WQ9735023-A2.

25-SEP-1997.

19-MAR-1997: 97WO-US04409.

PR 20-MAR-1996; 96US-0618911.

PA (PTON-) PIONEER HI-BRED INT. INC.

PT Coughlan S, Hastings C, Hu D, Jung R;

WP: 1997-480228/44.

DR N-PSDB: AAT86283.

AA Increasing the nutritional quality of soybean - by genetically
PT increasing the amount of lysine, methionine and/or cysteine in
PT albumin proteins
PT

xx
ps
Claim 25: page 48-49; 62pp; English.

This sequence represents a chimeric soybean albumin protein. This sequence was created from the albumin 1 and albumin 3 protein sequences (see AAW23418 and AAW23419). The DNA encoding this sequence represents a DNA molecule of the invention. The DNA molecules of the invention comprise a preselcted DNA segment encoding a seed storage protein (SSP). It is used to produce transgenic seeds and plants, especially soybean plants that have increased levels of lysine and also increased levels of methionine and/or cysteine in albumin protein type 1 and/or 3. Increasing the levels of these amino acids increases the nutritional value of soybean produced. This is especially useful for producing animal feeds. The amount of lysine in the seed is increased by 5-10 %, the amount of methionine and cysteine is increased by 10-30 %

XX	Sequence	158 AA:
SO		

Query Match	21.5%;	Score 239;	DB 18;	Length 158;
Best Local Similarity	35.4%;	Pred. No. 1.8e-17;		
Matches	56: Conservative	32: Mismatches	58: Indels	12: Gaps

1	QY	MAKUTILVALAFLAAHAASAROQWELQGRQCQSOLERANLRFCEOHLWQKIQ--RDED	50
1	Db	mtkftillilisilfci-ahtcsaskwqhgd-scrkqlkgvnltpcekhimekiqgrgdd	58
59	QY	SVYERDPSPSDPSPSPYDRGAGSSQHQE-----RCCNELNEFFENNRCMCEALQOI	112
59	Db	dddddnhilmtgrinyrkkegdedeeggmqccctemselk-spkcgcaklqi	117
113	QY	MENOSDRLGQQSQQFKRELRLNPQCCGLRAPQRCDL	150
118	Db	meagseele-ekenkkmekelmplmtmc:fgomigacd	154

RESULT 15
AAW23586
ID AAW23586 standard: protein: 158 AA:

XX AAW23586:

30-SEP-1997 (first entry)

XX Mahinlin MBLJ from Capparidaceae.

XX MBL; sweetener; transgenic plant; recombinant protein; food;
KW KW beverage; animal feed; chewing gum; dental hygiene product;
KW KW pharmaceutical.
KW

[illegible]



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 27, 2001, 13:03:11 ; Search time 8470.79 Seconds
(without alignments)
1328.220 Million cell updates/sec

Title: US-09-715-036-3

Perfect score: 682

Sequence: 1 gacacagaccactggtaat.....ctgtgactagcggtga 682

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl :

- 1: gb_ba.*
- 2: gb_htg.*
- 3: gb_in.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vi.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_om.*
- 20: em_or.*
- 21: em_ov.*
- 22: em_pat.*
- 23: em_ph.*
- 24: em_pl.*
- 25: em_ro.*
- 26: em_sts.*
- 27: em_sy.*
- 28: em_un.*
- 29: em_vi.*
- 30: em_htgo_hum.*
- 31: em_htgo_inv.*
- 32: em_htgo_rod.*
- 33: em_htg_hum.*
- 34: em_htg_inv.*
- 35: em_htg_rod.*
- 36: em_htg_other.*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No. Query Match Length DB ID

1	682	100.0	682	6	AX148740	AX148740 Sequence
2	682	100.0	1162	6	AX148738	AX148738 Sequence
3	462	67.7	717	8	ARQALIII	L77197 Arachis hyp
4	459.4	67.4	474	6	AX155333	AX155333 Sequence
5	157.4	23.1	627	8	AF092846	AF092846 Arachis h
6	149.2	21.9	5787	12	PEX2	X03174 Cloning vec
7	149.2	21.9	5787	12	SYNPEX2V	L08868 PEX2 expres
8	149.2	21.9	6728	12	PUEX	L06291 Bacterial e
9	146.8	21.5	1284	6	I00684	I00684 Sequence 12
10	146.8	21.5	3234	6	Al1623	Al1623 59 Val-IGF-
11	146.8	21.5	3234	6	Al1624	Al1624 59 Val-IGF-
12	146.8	21.5	3234	6	E00609	E00609 DNA encodin
13	146.4	21.5	12294	12	DMCASPLCZ	X81643 D.melanogas
14	146	21.4	350	10	S89402S1	S89402 ferritin Lg
15	146	21.4	1989	12	CVSACKPN	X16973 Cloning vec
16	146	21.4	3078	1	ECLACZ	V00296 E. coli gen
17	146	21.4	3150	6	AX163849	AX163849 Sequence
18	146	21.4	3152	6	A20699	A20699 pMC LacZ DN
19	146	21.4	3152	6	I15648	I15648 Sequence 3
20	146	21.4	3365	6	I18302	I18302 Sequence 1
21	146	21.4	3365	6	I72358	I72358 Sequence 1
22	146	21.4	3575	12	SYNPRBGA	M81126 Artificial
23	146	21.4	3624	12	SYNBPRA	M81125 Artificial
24	146	21.4	3652	6	AX114850	AX114850 Sequence
25	146	21.4	3652	6	AX137076	AX137076 Sequence
26	146	21.4	4342	12	SCU54830	U54830 Synthetic E
27	146	21.4	5019	6	Al1178	Al1178 Synthetic n
28	146	21.4	5019	6	E01007	E01007 DNA sequenc
29	146	21.4	5346	12	SYNPUR288V	L09147 PUR288 clon
30	146	21.4	5347	12	SYNPUR289V	L09148 PUR289 clon
31	146	21.4	5348	12	SYNPUR291V	L09150 PUR291 clon
32	146	21.4	5349	12	SYNPUR292V	L09151 PUR292 clon
33	146	21.4	5351	12	SYNPUR278V	L09146 PUR278 clon
34	146	21.4	5353	12	SYNPUR290V	L09149 PUR290 clon
35	146	21.4	5730	12	SYNLACZD	M34299 Cloning vec
36	146	21.4	5808	12	SYNLACZL	M34307 Cloning vec
37	146	21.4	5932	12	U03993	U03993 Cloning vec
38	146	21.4	6096	12	SYNLAC2B	M34297 Cloning vec
39	146	21.4	6107	12	SCU54829	U54829 Synthetic E
40	146	21.4	6114	12	CVU46489	U46489 Cloning vec
41	146	21.4	6119	12	SCU54828	U54828 Synthetic E
42	146	21.4	6168	12	CVU46490	U46490 Cloning vec
43	146	21.4	6252	12	CVU46491	U46491 Cloning vec
44	146	21.4	6280	12	XXU19930	U19930 Cloning vec
45	146	21.4	6476	12	SYNLACZAB	M34296 Cloning vec

ALIGNMENTS

RESULT 1

AX148740
LOCUS AX148740 682 bp DNA
DEFINITION Sequence 3 from Patent WO0136621.
ACCESSION AX148740
VERSION AX148740.1 GI:14347294
KEYWORDS
SOURCE peanut.
ORGANISM Arachis hypogaea

PAT 08-JUN-2001

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae;
Aschynomeneae; Arachis.
1 (bases 1 to 682)
Dodo,H.W., Arnzen,C.J., Konan,K.N. and Viquez,O.M.
Down-regulation and silencing of allergen genes in transgenic
peanut seeds
Patent: WO 0136621-A 3 25-MAY-2001;
Alabama A & M University (US)
Location/Qualifiers

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

[illegible]

Db 712 TGCTGTTGACTGTAGCGCTGA 733

RESULT 3
LOCUS ARQALLII 717 bp DNA 28-MAR-1996
DEFINITION Arachis hypogaea (clone Ara h II p38) allergen II gene, polyA signal.
ACCESSION L77197
VERSION L77197.1 GI:1336995
KEYWORDS allergen; conglutin; seed storage protein.
SOURCE Arachis hypogaea (strain Florunner) (clone: Ara h II p38) DNA.
ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Aeschynomeneae; Arachis.

REFERENCE 1 (bases 1 to 717)
AUTHORS Stanley, J.S.
TITLE The major peanut allergen Ara h II is a seed storage protein with multiple IgE-binding epitopes
JOURNAL Unpublished
FEATURES
source Location/Qualifiers
1. 717
/organism="Arachis hypogaea"
/strain="Florunner"
/db_xref="taxon:3818"
/clone="Ara h II p38"
polyA_signal 562..567
BASE COUNT 217 a 152 c 184 g 164 t
ORIGIN

Query Match 67.7%; Score 462; DB 8; Length 717;
Best Local Similarity 100.0%; Pred. No. 9.9e-113;
Matches 462; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 67 ctcaccatactagtagccctgcgccttttccctcctgcgtgccacgcacatctgcgagga 126
Db 1 GCTCACCATACTAGTAGCCCTGCGCCCTTTTCCTCTGCTGCCACGCATCTCGGAGGA 60

QY 127 gcagtggaactccaagagagacgaagatgccagagccagctgcagagggcggaacctgag 186
Db 61 GCAGTGGGAACCTCCAGGAGACAGAGATGCCAGAGCCAGCTCGAGAGGGCGCAACTGAG 120

QY 187 gccctgcagcaacatctcagcaagatcccaactgacagagattcattatgaacggga 246
Db 121 GCCCTGCGACACATCTCATGTGAGAGATCCCAACGTGACGAGGATTCATATGAACGGGA 180

QY 247 cccgtacagccctagtcaggatccgtacagccctagtcacatagtcagagagggcgtgg 306
Db 181 CCGGTACAGCCCTAGTCAGGATCCGTACAGCCCTAGTCATATATGATCGAGAGGGCGCTGG 240

QY 307 atcctctcagcagcagagagaggtgttgcaatgagctgaacaggtttgagaacacaaag 366
Db 241 ATCCTCTCAGCACCAGAGAGAGGTGTGCAATGAGCTGAACAGAGTTTGAGAACCAAG 300

QY 367 gtgcattgcgagcagcattgcaacagatcatggaacacagcagatagttgcaggggag 426
Db 301 GTGCATGTGCGAGGCATTCGAACAGATCATGTGAGAACCAAGAGGATAGTTGCAGGGGAG 360

QY 427 gcaacaggagcaacagtcgaagggagctcaggaacttgccctcaacagtcgagccttag 486
Db 361 GCAACAGGAGCAACAGTTCAAGAGGAGCTCAGGAACTTGCCTCAACAGTCGGCGCTTAG 420

QY 487 ggcacacagcgttcgacttgacatgcgaacagtgcaaaagtggcgag 528
Db 421 GGCACACACAGCGTTGGGACTTGGACGTCGAAAGTGCGGCGAG 462

RESULT 4
LOCUS AX155333 474 bp DNA PAT 22-JUN-2001

DEFINITION Sequence 6 from Patent WO0140264.
ACCESSION AX155333
VERSION AX155333.1 GI:14536768
KEYWORDS peanut.
SOURCE Arachis hypogaea
ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Aeschynomeneae; Arachis.

REFERENCE 1 (bases 1 to 474)
AUTHORS Bannon, G.A., Burks, W.A., Caplan, M.J., Sampson, H. and Sosin, H.
TITLE Peptide antigens
JOURNAL Patent: WO 0140264-A 6 07-JUN-2001;
Panacea Pharmaceuticals, LLC (US); The University of Arkansas (US); MOUNT SINAI SCHOOL OF MEDICINE OF THE CITY UNIVERSITY OF NEW YORK (US)

FEATURES
source Location/Qualifiers
1. 474
/organism="Arachis hypogaea"
/db_xref="taxon:3818"
BASE COUNT 132 a 126 c 138 g 78 t
ORIGIN

Query Match 67.4%; Score 459.4; DB 6; Length 474;
Best Local Similarity 99.8%; Pred. No. 5.1e-112;
Matches 460; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 68 ctcaccatactagtagccctgcgccttttccctcctgcgtgccacgcacatctgcgagga 127
Db 1 CTCACATACATAGTAGCCCTGCGCCCTTTTCCTCTGCTGCCACGCATCTCGGAGGCGAG 60

QY 128 cagtggaactccaagagagacgaagatgccagagccagctgcagagggcggaacctgag 187
Db 61 CAGTGGGAACCTCCAGGAGACAGAGATGCCAGAGCCAGCTCGAGAGGGCGCAACTGAGG 120

QY 188 cccctgcgagcaacatctcagcagaagatcccaactgacagagattcattatgaacgggac 247
Db 121 CCTCTCAGCACCAGAGAGAGGTGTGCAATGAGCTGAACAGAGTTTGAGAACCAAGG 180

QY 248 ccgtacagccctagtcaggatccgtacagccctagtcacatagtcagagagggcgctgga 307
Db 181 CCGTACAGCCCTAGTCAGGATCCGTACAGCCCTAGTCATATGATCGAGAGGGCGCTGGA 240

QY 308 tccctcagcagcagagagaggtgttgcaatgagctgaacaggtttgagaacacaaagg 367
Db 241 TCCTCTCAGCACCAGAGAGAGGTGTGCAATGAGCTGAACAGAGTTTGAGAACCAAGG 300

QY 368 tgcattgcgagcagcattgcaacagatcatggaacacagcagatagttgcaggggag 427
Db 301 TCATGTGCGAGGCATTCGAACAGATCATGTGAGAACCAAGAGGATAGTTGCAGGGGAGG 360

QY 428 caacaggagcaacagtcgaagggagctcaggaacttgccctcaacagtcgagccttagg 487
Db 361 CAACAGGAGCAACAGTTCAAGAGGAGCTCAGGAACTTGCCTCAACAGTCGGCGCTTAGG 420

QY 488 gcaccacagcgttcgacttgacatgcgaacagtgcaaaagtggcgag 528
Db 421 GCACCACAGCGTTGGGACTTGGACGTCGAAAGTGCGGCGAG 461

RESULT 5
LOCUS AF092846 627 bp mRNA PLN
DEFINITION Arachis hypogaea allergen Ara h 6 (Ara h 6) mRNA, partial cds.
ACCESSION AF092846
VERSION AF092846.1 GI:5923741
KEYWORDS peanut.
SOURCE Arachis hypogaea
ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;


```

artificial sequence.
1 (bases 1 to 5787)
AUTHORS
  Gilbert W.
TITLE
  Obtained from VecBase 3.0
JOURNAL
  Unpublished (1991)
COMMENT
  These data and their annotation were supplied to GenBank by Will
  Gilbert under the auspices of the GenBank Curator Program. pEX2 -
  Expression Vector
ENTRY pEX2 #TYPE DNA CIRCULAR
TITLE pEX2 - Expression Vector
DATE 27-MAR-1987
#sequence 12-MAY-1987
ACCESSION V0107
SOURCE artificial
REFERENCE
  #number 1
  #author Stanley K.K., Luzzio J.P.
  #journal EMBO J. (1984) 3: 1429-1434
  #title
    Construction of a new family of high efficiency bacterial
    vectors: identification of cDNA clones coding for human liver
    proteins
  #comment bases 1-5787; enum. 1 to 5786
REFERENCE
  #number 2
  #authors Pouwels P.H., Enger-Valk B.E., Brammar W.J.
  #book Cloning Vectors, Elsevier 1985 and supplements
  #comment vector 1-B-iv-5
COMMENT
  Fetched from the EMBL database, entry AREX2
  Revised 12-MAY-1987 by F. Pfeiffer to include revision to
  pBR322.
COMMENT from EMBL AREX2
Key From To Description
PRM 1 61 P(R) promoter from phage lambda
SITE 67 67 transcription start site
SITE 85 85 cro amino terminus
SITE 111 135 oligonucleotide
SITE 136 249 lacI fragment
SITE 250 3195 lacZ (aa 24-1005)
SITE 3197 3200 AAT deleted in pEX1
SITE 3200 3201 AATT inserted in pEX3
SITE 3196 3231 cloning linker
SITE 3234 3244 stop codon region
SITE 3196 3255 CGG missing in [1]
REVISION 3255 3256 GA inserted in [1]
REVISION 3263 3264 transcription terminator 1 from
SITE 3269 3371
phage fd
SITE 3378 3480 transcription terminator 2 KEYWORDS
CROSSREFERENCE
  #prevised
  #EMBL(11):AREX2
  #parent
  VecBase(3):pBR322, VecSource(3):Poly8, GenBank(50):lam,
  GenBank(50):PFD,
  GenBank(50):Ecolac, VecBase(3):pKTH601
  #brother
  VecBase(3):pEX1, VecBase(3):pEX3
PARENT
  Features pf pEX2 (5787 bp)
    residue source
    1- 110 37957-38066 phage lambda
    135- 251 1029- 1145 Ecolac
    249-3202 1355- 4308 Ecolac
    3196-3231 1- 36 M13mp8/pUC8-polylinker
    3234-3244 1- 11 STOP-oligonucleotide
    3269-3372 1500- 1603 phage fd
    3378-3481 1500- 1603 phage fd
    3487-5781 2069- 4363 pBR322
    Conflict (cfl) and Mutations (mut):
      pEX2 source
      cfl 3387 A C 1509 phage fd

```

```

mut 5029 A G 3611 pBR322
mut 5330 T C 3912 pBR322
FEATURES
  4716-5504 789-1 (c) Ap-R; b-lactamase
  POLYLINKER EcoRI-SmaI-BamHI-SalI-PstI-HindIII
  SELECTION #resistance Ap
  SUMMARY pEX2 #length 5787 #checksum 4444.
  Location/Qualifiers
    1..5787
    /organism="synthetic construct"
    /db_xref="taxon:32630"
BASE COUNT 1382 a 1511 c 1541 g 1353 t
ORIGIN
Query Match 21.9%; Score 149.2; DB 12; Length 5787;
Best Local Similarity 95.1%; Pred. No. 2.7e-29;
Matches 154; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
QY 521 ggcgcagcggcgccgaattccgccgatactgacgggctccaggagtcgagcaccacaa 580
Db 3217 gtcgacggatcccgccgggaattccgccgatactgacgggctccaggagtcgagcaccacaa 3158
QY 581 tcccatatggaaccgcgcgatattcagccatgtgcctcttcgcgctgcagcagatggc 640
Db 3157 tccccatattggaaccgcgcgatattcagccatgtgcctcttcgcgctgcagcagatggc 640
QY 641 gatgctggtttccatccatcagttgctgtgactgtgacgtgacggcgta 682
Db 3097 gatgctggtttccatccatcagttgctgtgactgtgacggcgta 3056
RESULT 8
PUEX/c PUEX
LOCUS Bacterial expression vector pUEX2 with universal host specificity.
DEFINITION X06291
ACCESSION X06291
VERSION X06291.1 GI:58279
KEYWORDS Cloning vector; plasmid.
SOURCE Cloning vector pUEX2.
ORGANISM Cloning vector pUEX2
REFERENCE 1 (bases 1 to 6728)
AUTHORS Stanley,K.K.
TITLE Direct Submission
JOURNAL Submitted (23-NOV-1987) Stanley K., EMBL, 6900 Heidelberg,
Meyerohofstr. 1
REFERENCE 2 (bases 1 to 6728)
AUTHORS Bressan,G.M. and Stanley,K.K.
TITLE pUEX, a bacterial expression vector related to pEX with universal
host specificity
JOURNAL Nucleic Acids Res. 15 (23), 10056 (1987)
MEDLINE 88096502
FEATURES
  Location/Qualifiers
    source
    1.. 6728
    /organism="Cloning vector pUEX2"
    /db_xref="taxon:31852"
    935
    /note="transcription start site from P(RM)"
  misc_feature
    935
    /note="cr857 amino terminus"
  misc_feature
    946..962
    /note="O(R)3"
  misc_feature
    969..985
    /note="O(R)2"
  misc_feature
    993..1009
    /note="O(R)1"
  misc_feature
    1018
    /note="transcription start site from P(R)"
  misc_feature
    1036
    /note="cro amino terminus"
  misc_feature
    1081..1194
    /note="lacI fragment"

```

```
misc_feature 1195..4140
/note="lacZ fragment"
misc_feature 4141..4176
/note="cloning linker"
misc_feature 4142..4145
/note="AATT" deleted in pUXE1"
misc_feature 4145..4146
/note="AATT" inserted in pUXE3"
misc_feature 4179..4189
/note="stop codon region"
misc_feature 4214..4316
/note="transcription terminator 1 from phage fd"
misc_feature 4323..4425
/note="transcription terminator 2 from phage"
BASE COUNT 1629 a 1735 c 1727 g 1637 t
ORIGIN
Query Match 21.9%; Score 149.2; DB 12; Length 6728;
Best Local Similarity 95.1%; Pred. No. 2.7e-29;
Matches 154; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
Qy 521 ggcgcagcggcgccgaattccgcgcatactgacgggtccaggagtcgtgcgccacaa 580
Db 4162 gtcgacggatccccgggaattccggccgatactgacgggctccaggagtcgtgcgccacaa 4103
Qy 581 tcccatatgaaacccgtcgatattcgccatgtgccttctccgcgtgcagcagatgac 640
Db 4102 tcccatatgaaacccgtcgatattcgccatgtgccttctccgcgtgcagcagatgac 4043
Qy 641 gatgctggtttccatcagttgctgtgactgttagcgctga 682
Db 4042 gatgctggtttccatcagttgctgtgactgttagcgctga 4001

RESULT 9
100684/c
LOCUS 100684 1284 bp ss-DNA PAT 21-MAY-1993
DEFINITION Sequence 12 from Patent US 4745179.
ACCESSION I00684
VERSION I00684.1 GI:269096
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS Ueda, I., Niwa, M., Saito, Y., Sato, S., Ono, H. and Kitaguchi, T.
TITLE sup.59 Valine insulin-like growth factor I and process for production thereof
JOURNAL
Fujisawa Pharmaceutical Co., Ltd.;
Osaka, ;
JP;
FEATURES
source Location/Qualifiers
1..1284
/organism="unknown"
BASE COUNT 284 a 360 c 372 g 268 t
ORIGIN
Query Match 21.5%; Score 146.8; DB 6; Length 1284;
Best Local Similarity 98.7%; Pred. No. 1.4e-28;
Matches 148; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 533 ccgcgaattccgcgcatactgacgggtccaggagtcgtgcgccacaaatccccatagga 592
Db 1075 CCATGAATTCGGCGGATCTACGAGGCTCCAGGAGTCGTCGCCACCAATCCCATATGGA 1016
Qy 593 aacgcgtgatttagcattgctcttccgcgtgcagcagatgacgctggttt 652
Db 1015 AACCGTCGATATTACGCCATGTGCTTCTTCCGCGTGCACGATGGCGATGGCTGTTT 956
Qy 653 ccatcagttgctgtgactgttagcgctga 682
```

```
|||||
Db 955 CCATCAGTTGCTGTGACTGTAGCGGCTGA 926
|||||

RESULT 10
11623/c
LOCUS 11623 3234 bp DNA PAT 17-NOV-1993
DEFINITION 59 Val-IGF-I fused with beta galactosidase.
ACCESSION 11623
VERSION 11623.1 GI:491150
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS Ueda, I., Niwa, M., Saito, Y., Sato, S., Ono, H. and Kitaguchi, T.
TITLE 59 Valine insulin-like growth factor I and process for production thereof
JOURNAL
Fujisawa Pharmaceutical Co., Ltd
FEATURES
source
1..3234
/organism="synthetic construct"
/db_xref="taxon:32630"
1..3234
/gene="59val-IGF-I fused with beta galactosidase"
1..3234
/gene="59val-IGF-I fused with beta galactosidase"
/notes="Protein sequence is in conflict with the conceptual translation"
/codon_start=1
/transl_table=11
/protein_id="CAA00972.1"
/db_xref="GI:4526654"
/translation="MTDLSAVLQRRDWPNGVTLNRLAAHPFASWRNSEART
DRPSQQLRSINGEWRFPAPPEAPESWLECDLPADTVVVPNSNMHGVDPIYTN
VTPYITVNPFPVPTENPTGCSYSLTFNVDENWLOEGOTRIIFDGVNSAFHLACNGRWG
YGQDSRLPSEFDLSAPLRAGENRLAVMLVRLWSDSGSYLEDQDMWRMSGIFRDSYLLHKP
TQISDFHVAFTNDFDSRAVLEAEVQMGCELDRLYRVTVSLMOWGTADGTLEAEACDVGFR
EIIDERGAYADRVTLRLNENKPLWSAEINLYRAVELHTADGTLEAEACDVGFR
VRIENGLLLKNGKPLLIRGVNREHHPHGLHQVMDEQTMVQDILLKQNNFNARCSHY
PNHLYTLCDRYGLYVVDENIETHGKVPNMRLTDDPRLPAMSERVTRMFAVRDRNH
PSVLIWSLGNESGHGANHDALYRWIKSVDPSPVOYEGGADTTATDIIICPMYARVDE
DQFPAPPKWSIKKWLSPGCTEPLILCEYAHAMGNSLGGFPAKYWQAFROYPLRQGGF
VNDWDQSLIKYDENGPNHSAVGGDFGPTDNDQFCMNGLYPADRTPHPALTEAKHQQ
QFFQRLSGTIEVTSYELFRSDNELLHWMVALDGKPLASGEFLDVPAPGKQLIEL
PELPQPSAGQLWLTVRVVPNPATAWSAGHISAWQWRLAENLSVTLPAPASHAIPHL
TTSEMDFCIELGNKRWQFNROSGFLSQMMIGDKKOLLTLPLRQDFTRAPLDNDIGVSEA
TRIDPNAWVERKKAAGHYQAEAAALQCCTADTLADAVLITTAHAWHQCKTLFISRKTY
RIDGSGOMALTVDVEVASDTPHPRIGLNCOLAQVAERVNWLGICPOENYPDRLTAAAC
FDRMDLPLSDMYTPVVPSENGLRGSTRRELANYGPHQWRGDFQFNISRYSOQLMETSH
RHLHAEETWLNIDGFHNGIGDSDSWSPSVAEPMGPETLCGRELVDALQFVCGDRG
FYFNKPTGYGSSRRAPQIGIVDECCFRSCDLRLRLEVTCAPLKPAKSA"
BASE COUNT 705 a 886 c 929 g 714 t
ORIGIN
Query Match 21.5%; Score 146.8; DB 6; Length 3234;
Best Local Similarity 98.7%; Pred. No. 1.3e-28;
Matches 148; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 533 ccgcgaattccgcgcatactgacgggtccaggagtcgtgcgccacaaatccccatagga 592
Db 3025 CCATGAATTCGGCGGATCTACGAGGCTCCAGGAGTCGTCGCCACCAATCCCATATGGA 2966
Qy 593 aacgcgtgatttagcattgctcttccgcgtgcagcagatgacgctggttt 652
Db 2965 AACCGTCGATATTACGCCATGTGCTTCTTCCGCGTGCACGATGGCGATGGCTGTTT 2906
Qy 653 ccatcagttgctgtgactgttagcgctga 682
Db 2905 CCATCAGTTGCTGTGACTGTAGCGGCTGA 2876
```

```
RESULT 11
A11624
LOCUS       A11624      3234 bp      DNA              PAT      17-NOV-1993
DEFINITION   59 Val-IGF-I fused with beta galactosidase.
ACCESSION   A11624
VERSION     A11624.1  GI:489368
KEYWORDS
SOURCE      synthetic construct.
ORGANISM    artificial sequence.
REFERENCE   1 (bases 1 to 3234)
AUTHORS     Ueda,I., Niwa,M., Saito,Y., Sato,S., Ono,H. and Kitaguchi,T.
TITLE       59 Valine insulin-like growth factor I and process for production thereof
JOURNAL
PATENT      EP 0158892-A 120 23-OCT-1985;
FUJISAWA PHARMACEUTICAL CO., LTD
FEATURES
source      Location/Qualifiers
            1..3234
            /organism="synthetic construct"
            /db_xref="taxon:32630"
BASE COUNT  714 a  929 c  886 g  705 t
ORIGIN

Query Match      21.5%; Score 146.8; DB 6; Length 3234;
Best Local Similarity 98.7%; Pred. No. 1.3e-28;
Matches 148; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 533 ccgcgaattccgcgatactgacgggtccagagtcgctgcgccaccaatccccatatgga 592
|| |||||
DB 210 CCATGAATTCGGCGGATACCTACGGGCTCCAGGAGTCGTCGCCACCAATCCCATATGGA 269
|||||
QY 593 aacgcgtgatccagccatgtcctttctccgcgtgcagcagatggcgatggctggttt 652
|||||
DB 270 AACCGTCGATATTACGCATGTGCTCTTCGCGCTGCAGCAGATGGCGATGGCTGGTTT 329
|||||
QY 653 ccatacgtgtctgtgactgtagcggctga 682
|||||
DB 330 CCATCAGTTCCTGTTGACTGTAGCGGCTGA 359
|||||

RESULT 12
E00609/c
LOCUS       E00609      3234 bp      DNA              PAT      29-SEP-1997
DEFINITION   DNA encoding (beta-gal)-(59 Val-IGFI) fused protein.
ACCESSION   E00609
VERSION     E00609.1  GI:2168888
KEYWORDS    JP 1986001397-A/3.
SOURCE      synthetic construct.
ORGANISM    artificial sequence.
REFERENCE   1 (bases 1 to 3234)
AUTHORS     Ueda,I., Niwa,M., Saito,Y., Sato,S., Ono,H. and Kitaguchi,T.
TITLE       GROWTH FACTOR I-LIKE 59VALINEINSULIN AND ITS PREPARATION
JOURNAL     Patent: JP 1986001397-A 3 07-JAN-1986;
FUJISAWA PHARMACEUT CO LTD
COMMENT     OS Artificial gene
            PC Artificial sequence; Genes.
            PN JP 1986001397-A/3
            PD 07-JAN-1986
            PF 01-APR-1985 JP 1985069630
            PR 02-APR-1984 GB 84 8408473, 01-JUN-1984 GB 84 8413989, PR
            25-SEP-1984 GB 84 8424157
            PI UEDA IKUO, NIWA MINEO, SAITO YOSHIMASA, SATO SUSUMU, PI ONO
            HIROKI,
            PI KITAGUCHI TADASHI
            PC C12P21/02,A61K35/74,A61K37/24,A61K37/26,C07H21/04,C12N1/00, PC
            C12N15/00,
            PC (C12P21/02,C12R1:19),(C12N1/00,C12R1:19);
            CC strandedness: Double;
            CC topology: Linear;
            CC hypothetical: No;
```

```
CC anti-sense: No; Location/Qualifiers
FH Key
FH CDS <1..>3234
FT /product="(beta-gal)-(59 Val-IGFI)fused FT
FT protein,
FT misc_feature 1..3021
FT /note="beta-gal coding region" FT
FT misc_feature 3025..3234
FT /note="59 Val-IGFI coding region".
FEATURES
source      Location/Qualifiers
            1..3234
            /organism="synthetic construct"
            /db_xref="taxon:32630"
BASE COUNT  705 a  889 c  926 g  714 t
ORIGIN

Query Match      21.5%; Score 146.8; DB 6; Length 3234;
Best Local Similarity 98.7%; Pred. No. 1.3e-28;
Matches 148; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 533 ccgcgaattccgcgatactgacgggtccagagtcgctgcgccaccaatccccatatgga 592
|| |||||
DB 3025 CCATGAATTCGGCGGATACCTACGGGCTCCAGGAGTCGTCGCCACCAATCCCATATGGA 2966
|||||
QY 593 aacgcgtgatccagccatgtcctttctccgcgtgcagcagatggcgatggctggttt 652
|||||
DB 2965 AACCGTCGATATTACGCATGTGCTCTTCGCGCTGCAGCAGATGGCGATGGCTGGTTT 2906
|||||
QY 653 ccatacgtgtctgtgactgtagcggctga 682
|||||
DB 2905 CCATCAGTTCCTGTTGACTGTAGCGGCTGA 2876
|||||

RESULT 13
DMCASPLCZ/c
LOCUS       DMCASPLCZ  12294 bp      DNA              SYN      29-APR-2000
DEFINITION   D.melanogaster P element CasPer-hs43-lacZ gene transformation vector.
ACCESSION   X81643
VERSION     X81643.1  GI:551429
KEYWORDS    Carnegie-4 vector; P element; transformation vector; white gene.
SOURCE      synthetic construct.
ORGANISM    artificial sequence.
REFERENCE   1 (bases 1 to 12294)
AUTHORS     Pirrotta,V. and Zeng,C.
JOURNAL     Unpublished
REFERENCE   2 (bases 1 to 12294)
AUTHORS     Pirrotta,V.
TITLE       Direct Submission
JOURNAL     Submitted (14-SEP-1994) V. Pirrotta, Dept of Zoology, University of Geneva, 30 Quai E. Ansermet, 1211 Geneva, SWITZERLAND
COMMENT     Related sequences: M28731, M30841, X02974.
FEATURES
source      Location/Qualifiers
            1..12294
            /organism="synthetic construct"
            /db_xref="taxon:32630"
            /focus
            source      1..12294
            /organism="Drosophila melanogaster"
            /db_xref="taxon:7227"
            source      1..12294
            /organism="Escherichia coli"
            /db_xref="taxon:562"
            misc_feature 1..104
            /note="polylinker"
            promoter 111..245
            /gene="hs43"
            source order(111..245,290..417,4625..5366,7396..8208,8220..12294)
            /organism="Drosophila melanogaster"
            /db_xref="taxon:7227"
```

gene	111.. 245	GLINQADVEPGEISCTSSNTCPSSGKVILETLNFSAADPLDYVGLAILIVSRV
mRNA	/gene="hs43" 154.. >3467 /gene="adh-lacZ-SV40 gene construct (290.. 4598)" /note="D.melanogaster adh gene leader (290.. 417); E.coli lacZ gene (423.. 3729)"	LAYLALRLRARKK" complement(9542.. 9611) /gene="white" /number=5 complement(9612.. 9743) /gene="white" /number=5 complement(9744.. 9963) /gene="white" /number=4 complement(9964.. 10309) /gene="white" /number=4 complement(10310.. 10341) /gene="white" /number=3 complement(10342.. 10996) /gene="white" /number=3 complement(10997.. 11070) /gene="white" /number=2 complement(11071.. 11342) /gene="white" /number=2 complement(11343.. 11737) /gene="white" /number=1 complement(11738.. 11984) /gene="white" /number=1
gene	154.. 3467	
CDS	327.. 3467 /gene="adh-lacZ-SV40 gene construct (290.. 4598)" /note="adh-lacZ-SV40 gene construct (290.. 4598)" /note="unnamed protein product" /codon_start=1 /transl_table=11 /protein_id="CAA57302.1" /db_xref="GI:551430"	
source	423.. 3729 /organism="Escherichia coli" /db_xref="taxon:562" 4625.. 4863 /note="p element 3' end" 4864.. 5366 /note="white linker sequences" 5367.. 7395 /note="puC8 sequences" 7396.. 7623 /note="white linker sequences" 7624.. 8208 /note="p element 5' end" complement(8796.. 11984) /gene="white" complement(8796.. 9541) /gene="white" /number=6 complement(join(8796.. 9541,9612.. 9743,9964.. 10309, 10342.. 10996,11071.. 11342,11738.. 11984)) /gene="white" complement(join(8927.. 9541,9612.. 9743,9964.. 10309, 10342.. 10996,11071.. 11342,11738.. 11808)) /gene="white" /codon_start=1 /transl_table=11 /protein_id="CAA57303.1" /db_xref="GI:870998"	BASE COUNT 3273 a 2916 c 3008 g 3097 t ORIGIN Query Match 21.5%; Score 146.4; DB 12; Length 12294; Best Local Similarity 93.3%; Pred. No. 1.4e-28; Matches 153; Conservative 0; Mismatches 11; Indels 0; Gaps 0; Qy 519 gtggcggcaggcggcgaattccgcgcgatactgcagggctccagagtcgctgcgcacc 578 Db 3437 GTAGGACCGCGCTCAGCTGCTCGCGGATCTGACGGGCTCCAGGAGTCTGCCACC 3378 Qy 579 aatccccataggaaacgcgcgataattcagccatgtgcttcttcgcgcgcacagatg 638 Db 3377 AATCCCCATATGGAACCGTCGATATTCAGCCATGTGCTTCTTCCGCGTCGACGATG 3318 Qy 639 gcgatggctgtttccatcatcagttgctgtgactgtagcggctga 682 Db 3317 GCGATGGCTGGTTTCCATCAGTTGCTGTGTGACTGTAGCGGCTGA 3274 RESULT 14 S89402S1/c LOCUS DEFINITION ferritin Ig subunit [5' and 3' regions] (mice, Genomic, 350 nt, segment 1 of 2). ACCESSION S89402 VERSION S89402.1 GI:247107 KEYWORDS SEGMENT i of 2 Mus sp. Mus sp. ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. REFERENCE 1 (bases 1 to 350) AUTHORS Renaudie F., Yachou A.K., Grandchamp B., Jones R. and Beaumont C. TITLE A second ferritin L subunit is encoded by an intronless gene in the mouse JOURNAL Mamm. Genome 2 (3), 143-149 (1992) MEDLINE 92182532 REMARK GenBank staff at the National Library of Medicine created this

entry [NCBI gibbsq 89402] from the original journal article.
This sequence comes from Figure 5.

FEATURES

source
1. .350
/organism="Mus sp."
/db_xref="taxon:10095"
BASE COUNT 85 a 99 c 88 g 77 t 1 others
ORIGIN

Query Match 21.4%; Score 146; DB 10; Length 350;
Best Local Similarity 100.0%; Pred. No. 2.6e-28;
Matches 146; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 537 gaattccgcgatactgacgggtccaggagtcgtccccaatacccccatatggaaacc 596
|||||
Db 265 GAATTCCGCCGATACTGACGGGCTCCAGGAGTCGTCGCCACCAATCCCATATGGAACC 206
|||||
QY 597 gtccatattcagccatgtgccttcttcgcgtcagcagatggcgatggctgtttccat 656
|||||
Db 205 GTCGATATTACGCATGTGCCTTCTTCCGCGTCGACGAGATGGCGATGGCTGTTTCCAT 146
|||||
QY 657 cagttgtgttgactgtagcggtga 682
|||||
Db 145 CAGTTGCTGTGACTGTAGCGGCTGA 120
|||||

RESULT 15

CVSACKPN/c CVSACKPN 1989 bp DNA SYN 15-NOV-1990
LOCUS Cloning vector lambda-gt11 DNA sequence of SacI-KpnI region.
DEFINITION
ACCESSION X16973
VERSION X16973.1 GI:58246
KEYWORDS Cloning vector.
SOURCE Cloning vector lambda gt11.
ORGANISM Cloning vector lambda gt11
artificial sequence; vectors.
REFERENCE 1 (bases 1 to 1989)
Slatko, B.E.
AUTHORS
TITLE Direct Submission
JOURNAL Submitted (20-OCT-1989) Slatko B. E., New England Biolabs, Inc., 32
Tozer Road Beverly MA 01915, USA
REFERENCE 2 (bases 1 to 1989)
Moran, L.S., Maina, C.V., Poole, C.B. and Slatko, B.E.
AUTHORS Nucleotide sequence of the phage lambda gt11 SacI-KpnI lacZ region
TITLE
JOURNAL Gene 93 (1), 163-164 (1990)
MEDLINE 91033010
COMMENT See <X04125> for lambda gt11 lac operon and phage junction. Data
kindly reviewed (15-DEC-1989) by Slatko B.

FEATURES

source
1. .1989
/organism="Cloning vector lambda gt11"
/db_xref="taxon:31786"
/clone="pBS:Luc8"
/clone_lib="rest frags SacI-KpnI double digest"
misc_feature 1. .6
/note="SacI restriction site"
misc_feature 1069. .1074
/note="EcoRI restriction site"
misc_feature 1984. .1989
/note="KpnI restriction site"
BASE COUNT 471 a 597 c 518 g 403 t
ORIGIN

Query Match 21.4%; Score 146; DB 12; Length 1989;
Best Local Similarity 100.0%; Pred. No. 2.2e-28;
Matches 146; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 537 gaattccgcgatactgacgggtccaggagtcgtccccaatacccccatatggaaacc 596
|||||
Db 1074 GAATTCCGCCGATACTGACGGGCTCCAGGAGTCGTCGCCACCAATCCCATATGGAACC 1015
|||||

QY 597 gtccatattcagccatgtgccttcttcgcgtcagcagatggcgatggctgtttccat 656
|||||
Db 1014 GTCGATATTACGCATGTGCCTTCTTCCGCGTCGACGAGATGGCGATGGCTGTTTCCAT 955
|||||
QY 657 cagttgtgttgactgtagcggtga 682
|||||
Db 954 CAGTTGCTGTGACTGTAGCGGCTGA 929
|||||

Search completed: December 27, 2001, 13:03:21
Job time: 13867 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 27, 2001, 13:14:47 ; Search time 486.97 Seconds
(without alignments)
1200.681 Million cell updates/sec

Title: US-09-715-036-3

Perfect score: 682

Sequence: 1 gacacagaccactggttaa.....ctgttgacttagcggtga 682

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

- N_Geneseq_1101.*
- 1: /SIDS2/gcgdata/geneseq/geneseq/NA1980.DAT.*
 - 2: /SIDS2/gcgdata/geneseq/geneseq/NA1981.DAT.*
 - 3: /SIDS2/gcgdata/geneseq/geneseq/NA1982.DAT.*
 - 4: /SIDS2/gcgdata/geneseq/geneseq/NA1983.DAT.*
 - 5: /SIDS2/gcgdata/geneseq/geneseq/NA1984.DAT.*
 - 6: /SIDS2/gcgdata/geneseq/geneseq/NA1985.DAT.*
 - 7: /SIDS2/gcgdata/geneseq/geneseq/NA1986.DAT.*
 - 8: /SIDS2/gcgdata/geneseq/geneseq/NA1987.DAT.*
 - 9: /SIDS2/gcgdata/geneseq/geneseq/NA1988.DAT.*
 - 10: /SIDS2/gcgdata/geneseq/geneseq/NA1989.DAT.*
 - 11: /SIDS2/gcgdata/geneseq/geneseq/NA1990.DAT.*
 - 12: /SIDS2/gcgdata/geneseq/geneseq/NA1991.DAT.*
 - 13: /SIDS2/gcgdata/geneseq/geneseq/NA1992.DAT.*
 - 14: /SIDS2/gcgdata/geneseq/geneseq/NA1993.DAT.*
 - 15: /SIDS2/gcgdata/geneseq/geneseq/NA1994.DAT.*
 - 16: /SIDS2/gcgdata/geneseq/geneseq/NA1995.DAT.*
 - 17: /SIDS2/gcgdata/geneseq/geneseq/NA1996.DAT.*
 - 18: /SIDS2/gcgdata/geneseq/geneseq/NA1997.DAT.*
 - 19: /SIDS2/gcgdata/geneseq/geneseq/NA1998.DAT.*
 - 20: /SIDS2/gcgdata/geneseq/geneseq/NA1999.DAT.*
 - 21: /SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT.*
 - 22: /SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
1	682	100.0	682	22	AA90337
2	682	100.0	1162	22	AA90336
3	462	67.7	717	18	AAT76614
4	462	67.7	717	18	AAT76615
5	461	67.6	474	20	AA06383
6	461	67.6	474	20	AA06388
7	146	21.4	3150	22	AAH45081
8	146	21.4	3365	14	AAQ41287
9	146	21.4	3652	22	AAD04924
10	146	21.4	3652	22	AAF85070
11	146	21.4	3756	20	AAH87978

C 12	146	21.4	4279	20	AA06800
C 13	146	21.4	4487	21	AAA09085
C 14	146	21.4	5201	21	AAA09087
C 15	146	21.4	5346	21	AAA09089
C 16	146	21.4	5835	21	AAA09091
C 17	146	21.4	6050	22	AAD10237
C 18	146	21.4	6560	20	AA026118
C 19	146	21.4	6561	20	AA026117
C 20	146	21.4	6700	22	AAH74523
C 21	146	21.4	6796	13	AAQ27433
C 22	146	21.4	6995	16	AAQ85466
C 23	146	21.4	7123	20	AAH86929
C 24	146	21.4	7174	10	AAH91062
C 25	146	21.4	7175	22	AAD04927
C 26	146	21.4	7252	22	AAH86928
C 27	146	21.4	7319	19	AAV60246
C 28	146	21.4	7573	22	AAD04930
C 29	146	21.4	7573	22	AAD04944
C 30	146	21.4	8062	22	AAD04946
C 31	146	21.4	8136	18	AAH63236
C 32	146	21.4	8153	22	AAD04945
C 33	146	21.4	8518	22	AAH74524
C 34	146	21.4	8574	20	AAH77158
C 35	146	21.4	8657	19	AAV33614
C 36	146	21.4	8710	16	AAQ89650
C 37	146	21.4	9020	16	AAT07652
C 38	146	21.4	9093	19	AAV33623
C 39	146	21.4	9093	20	AAH77157
C 40	146	21.4	9138	22	AAH83047
C 41	146	21.4	9146	20	AAH77156
C 42	146	21.4	9641	20	AAH15626
C 43	146	21.4	9729	22	AAH83093
C 44	146	21.4	9756	20	AAH11457
C 45	146	21.4	9756	20	AAH61062

ALIGNMENTS

RESULT 1

AA90337
ID AA90337 standard; DNA; 682 BP.

XX AA90337;

XX 23-JUL-2001 (first entry)

XX Peanut allergen Ara h2 gene.

XX Peanut; allergen; Ara h2; Ara h6; Ara h7; transgenic plant;

XX allergy; ds.

XX Arachis hypogaea.

XX Key Location/Qualifiers

FT CDS 59..682

FT sig_peptide /*tag= a

FT mat_peptide /*tag= b

FT misc_feature /*tag= c

FT /*tag= d

FT /*note= *PCR amplified fragment*

WO200136621-A2.

PD 25-MAY-2001.

XX 20-NOV-2000; 2000WO-US31657.

XX 19-NOV-1999; 99US-0167255.

SN on bar

X

Db 301 gtgcatgtgcaggcattcaacagatcatgagaaaccagagcagatggttcgaggag 360
 QY 427 gcaacaggagcaacagttcaaggaggagctcaggaaacttgctcaacagtgaggccttag 486
 Db 361 gcaacaggagcaacagttcaaggaggagctcaggaaacttgctcaacagtgaggccttag 420
 QY 487 gcaacacagcgttcgacttgacgtcagctcaaaagtgcgcag 528
 Db 421 ggcaccacagcgttcgacttgacgtcagctcaaaagtgcgcag 462

RESULT 4
 AAT76615
 ID AAT76615 standard; cDNA to mRNA; 717 BP.
 XX
 AC AAT76615;
 XX
 DT 29-DEC-1997 (first entry)
 XX
 DE Peanut allergen Ara hII cDNA clone.
 XX
 KW Peanut seed storage protein; allergen; allergy; hypersensitivity;
 KW vaccine; anaphylactic shock; immunotherapy; therapy;
 KW monoclonal antibody; ELISA; analysis; Ara hII; ds.
 XX
 OS Arachis hypogaea strain Florunner.
 XX

PH Key Location/Qualifiers
 FT CDS 2..475
 FT polyA_signal 562..567
 FT /*tag= a
 FT /*tag= b
 XX
 PN W09724139-A1.
 XX
 PD 10-JUL-1997.
 XX
 PF 23-SEP-1996; 96WO-US15222.
 XX
 PR 04-MAR-1996; 96US-0610424.
 PR 29-DEC-1995; 95US-0009455.
 XX
 PA (UYAR-) UNIV ARKANSAS.
 XX
 PI Bannon GA, Burks AW, Cockrell G, Helm RM, Stanley JS;
 XX
 DR WPI; 1997-363453/33.
 DR P-PSDB; AAW24153.

Peanut allergens Ara hI and Ara hII - used for vaccination and in two-site monoclonal antibody based ELISA
 XX
 PS Claim 31; Page 196; 354pp; English.
 XX

This cDNA clone codes for the major peanut allergen Ara hII (AAW22153), which contains multiple IgE-binding epitopes. It was amplified from a peanut seed cDNA library using a primer (see AAT58683) based on an isolated Ara hII peptide (see AAW24151). The sequence shows significant homology with the conglutin family of seed storage proteins of other legumes. The gene is capable of producing a protein product in prokaryotic cells that is recognised by serum IgE from a large proportion of individuals with peanut hypersensitivity. Ara hII and Ara hI (see AAW24149-50) can be used to raise monoclonal antibodies which are used in a specific two-site Mab ELISA for the detection of Ara hI or Ara hII (claimed). CC IgE-binding Ara hII antigen epitopes (see AAW24188-93) may be used CC in vaccines to protect against allergic reactions to peanut CC allergens, e.g. anaphylactic shock.
 XX

Sequence 717 BP; 217 A; 152 C; 184 G; 164 T; 0 other;
 SQ

Query Match 67.7%; Score 462; DB 18; Length 717;
 Best Local Similarity 100.0%; Pred. No. 3.4e-126;
 Matches 462; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 67 gctcaacatactagtagcccttcgccccttttctctctgctgcccacgcatctcgaggca 126
 Db 1 gctcaacatactagtagcccttcgccccttttctctctgctgcccacgcatctcgaggca 60
 QY 127 gcagtgggaactccaaggagacagaagatgccagagccagctcgagggcggaacctgag 186
 Db 61 gcagtgggaactccaaggagacagaagatgccagagccagctcgagggcggaacctgag 120
 QY 187 gccctgcgagcaacatctcatgcagaagatccaacgtgacgagattcatatgaacggga 246
 Db 121 gccctgcgagcaacatctcatgcagaagatccaacgtgacgagattcatatgaacggga 180
 QY 247 cccgtacagcccttagtcaggtacgacagcccttagtcctcatatgacggagggcgtgg 306
 Db 181 cccgtacagcccttagtcaggtacgacagcccttagtcctcatatgacggagggcgtgg 240
 QY 307 atcctctcagcaccagaagaggtgttgcaatgagctgaacgagtttgagaaacccaag 366
 Db 241 atcctctcagcaccagaagaggtgttgcaatgagctgaacgagtttgagaaacccaag 300
 QY 367 gtgcattgtcgaggcattgcacagatcatggaacacagagcagataggttcgagggag 426
 Db 301 gtgcattgtcgaggcattgcacagatcatggaacacagagcagataggttcgagggag 360
 QY 427 gcaacagagcaacagttcaaggaggaggtcaggaacttgcctcaacagtgccgcttag 486
 Db 361 gcaacagagcaacagttcaaggaggaggtcaggaacttgcctcaacagtgccgcttag 420
 QY 487 ggcaccacagcgttcgacttgacgtcgaaagtggcgagcag 528
 Db 421 ggcaccacagcgttcgacttgacgtcgaaagtggcgagcag 462

RESULT 5
 AAZ06383
 ID AAZ06383 standard; DNA; 474 BP.
 XX
 AC AAZ06383;
 XX
 DT 09-NOV-1999 (first entry)
 XX
 DE Peanut allergen, Ara h 2.
 XX
 KW allergy; immune response; transgenic; allergen; epitope;
 KW immunoglobulin E; Ig E; binding site; peanut; ds.
 XX
 OS Arachis hypogaea.
 XX
 PN W09538978-A1.
 XX
 PD 05-AUG-1999.
 XX
 PF 29-JAN-1999; 99WO-US02031.
 XX
 PR 27-AUG-1998; 98US-0141220.
 PR 31-JAN-1998; 98US-0073283.
 PR 13-FEB-1998; 98US-0074590.
 PR 13-FEB-1998; 98US-0074624.
 PR 13-FEB-1998; 98US-0074633.
 XX
 PA (SOSI/) SOSIN H.
 PA (UYAR-) UNIV ARKANSAS.
 PA (UYNI) UNIV NEW YORK WT SINAI SCHOOL MEDICINE.
 XX
 PI Bannon GA, Burks AW, Sampson HA, Sosin H;
 XX
 DR WPI; 1999-479189/40.
 DR P-PSDB; AAY15245.
 XX

AA7

QY 368 tgcattgagcagcattgcaacagatcatgagagaccagagatggtgcaggagg 427
 Db 301 tgcattgagcagcattgcaacagatcatgagagaccagagatggtgcaggagg 360
 QY 428 caacagagagcaacagttcaagagagagcagcaggaacttgcctcaacagtcgagccttagg 487
 Db 361 caacagagagcaacagttcaagagagagcagcaggaacttgcctcaacagtcgagccttagg 420
 QY 488 qcaccacagcgttgagcacttgagcgtcgaaagtgccgagcag 528
 Db 421 qcaccacagcgttgagcacttgagcgtcgaaagtgccgagcag 461

RESULT 7
 AAH45081/C
 ID AAH45081 standard; DNA; 3150 BP.
 XX
 AC AAH45081;
 XX
 DT 04-SEP-2001 (first entry)
 XX
 DE lacZ gene.
 XX
 KW lacZ; immunosuppressive; epigenetic regulation motif; immune response;
 KW T-cell response; methylation activity;
 KW methyl DNA binding protein identification; ds.
 XX
 OS Escherichia coli.
 XX
 PN W0200140478-A2.
 XX
 PD 07-JUN-2001.
 XX
 PF 06-DEC-2000; 2000WO-EP12793.
 XX
 PR 06-DEC-1999; 99CA-2291367.
 XX
 PA (INSP) INST PASTEUR.
 PA (CNRS) CENT NAT RECH SCI.
 XX
 PI Nicolas J, Henry I, Choulhka A;
 XX
 DR WPI; 2001-367812/38.
 XX

New isolated polynucleotide having reduced or increased content of epigenetic control motifs for studying, increasing and/or reducing gene expression, and improving DNA vaccination methods -
 Example 1; Fig 1; 75pp; English.
 The present invention relates to modified LacZ genes (see AAH45079 and AAH45080), in which epigenetic regulation motifs have been mutated compared to the wild-type sequence (the present sequence). The mutant LacZ genes are useful for inducing in a second host, a protective immune response, against a gene product of a first host. The mutant Lac Z genes are also useful for evaluating a promoter in biological systems, for comparing methylation activity in biologically systems and/or for identifying unknown methyl DNA binding proteins. The mutant lacZ genes are also useful for compensating a genetic defect, and for therapeutic applications. The mutant lacZ genes are also useful for minimising a T-cell response against the T-cells or tissues treated with them.
 Sequence 3150 BP; 706 A; 853 C; 904 G; 687 T; 0 other;

Query Match 21.4%; Score 146; DB 22; Length 3150;
 Best Local Similarity 100.0%; Pred. No. 5.7e-33;
 Matches 146; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 537 gaattccgcgcgatactgacggggtccaggagtcgctgcacacacatccccatggaacc 596
 Db 3099 GAATTCGCCGCGATACGCGGCTCCAGGAGTCGTCGCCACCAATCCCCATATGGAAACC 3040

QY 597 gtcgatattcagccatgtgcctttcttcgcgagatgagcagatggtggtttccat 656
 Db 3039 GTCGATATTCAAGCCATGTCCTTCTTCGCGTCGACGAGATGCGGTGTTCCAT 2980
 QY 657 cagttgctgttgactgttagcgctga 682
 Db 2979 CAGTTGCTGTTGACTGTAGCGGCTGA 2954

RESULT 8
 AAQ41287/C
 ID AAQ41287 standard; DNA; 3365 BP.
 XX
 AC AAQ41287;
 XX
 DT 26-AUG-1993 (first entry)
 XX
 DE Ubiquitin fusion protein, Ub-mat-beta-Gal, coding sequence.
 XX
 KW Fusion protein; ubiquitin-methionine-beta-galactosidase; UBP2; UBP3;
 KW expression vector; ubiquitin-specific; protease; UBPl; in vitro;
 KW transformation; Ub-met-beta-Gal; in vivo; ss.
 XX
 OS Synthetic.
 XX
 PN W09309235-A.
 XX
 PD 13-MAY-1993.
 XX
 PF 06-NOV-1992; 92WO-US09562.
 XX
 PR 08-NOV-1991; 91US-0789915.
 XX
 PA (MASI) MASSACHUSETTS INST TECHNOLOGY.
 XX
 PI Baker RT, Tobias JW, Varshavsky A;
 XX
 DR WPI; 1993-167700/20.
 DR P-PSDB; AAR36728.
 XX

Ubiquitin-specific protease(s) - specifically cleave fusion proteins regardless of size, at C-terminus of ubiquitin portion
 Claim 2; Page 31-36; 78pp; English.
 This sequence encodes the 120 kD fusion protein ubiquitin-methionine-beta-galactosidase (Ub-met-beta-Gal). This protein was used as a substrate to demonstrate the specific cleavage of the ubiquitin-specific proteases UBPl, UBP2, and UBP3 (see also AAR36729-31). The proteases UBPl and UBP2 demonstrate activity both in vivo and in vitro, whereas UBP3 is only active in vivo. These proteases have been shown to specifically cleave Ub-met-beta-Gal at the C-terminus of the ubiquitin moiety. The proteases can be used to deubiquitinate fusion proteins in vivo, therefore prokaryotic cells having an expression vector one of these proteases can be further transformed with an expression vector encoding a ubiquitin fusion protein. Such cells will then produce a deubiquitinated product having a predetermined N-terminal amino acid residue.
 Sequence 3365 BP; 765 A; 917 C; 951 G; 732 T; 0 other;

Query Match 21.4%; Score 146; DB 14; Length 3365;
 Best Local Similarity 100.0%; Pred. No. 5.9e-33;
 Matches 146; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 537 gaattccgcgcgatactgacggggtccaggagtcgctgcacacacatccccatggaacc 596
 Db 3315 GAATTCGCCGCGATACGCGGCTCCAGGAGTCGTCGCCACCAATCCCCATATGGAAACC 3256
 QY 597 gtcgatattcagccatgtgcctttcttcgcgagatgagcagatggtggtttccat 656
 Db 3255 GTCGATATTCAAGCCATGTCCTTCTTCGCGTCGACGAGATGCGGTGTTCCAT 3196

QY 657 cagttgctgttgactgtagcgctga 682
DB 3195 CAGTTGCTGTGACTGTAGCGGCTGA 3170

RESULT 9
AAD04924/c
ID AAD04924 standard; DNA; 3652 BP.

AC AAD04924;
XX
XX
DT 17-JUL-2001 (first entry)

XX E. coli beta-galactosidase reporter gene of gene trapping construct.
XX
XX Gene trapping construct; conditional mutation; unidirectional inversion;
KW recombinase recognition sequence; RRS; disruption cassette;
KW selection cassette; transgenic organism; beta-galactosidase; ds.
XX
XX Escherichia coli.

OS
XX
XX WO200129208-A1.

XX
XX
XX 26-APR-2001.

XX 16-OCT-2000; 2000WO-EP10162.

XX 16-OCT-1999; 99EP-0120592.

PR 27-OCT-1999; 99US-0162016.

XX
XX (ARTE-) ARTEMIS PHARM GMBH.

PA (FRAN-) FRANKEN BIOTECHNOLOGIE AG.

XX
XX Kuehn R, Von Melchener H, Altschmied J;

XX
XX WPI; 2001-308486/32.

XX New gene trapping construct capable of causing conditional mutations in
PT genes, comprises functional DNA segment inserted in sense or antisense
PT direction relative to gene to be trapped -
XX
XX Disclosure; Page 51-52; 78pp; English.

XX The present invention relates to a conditional gene trapping construct
CC capable of causing conditional mutations in genes. The gene trapping
CC construct comprises two functional DNA segments, each being flanked by
CC two recombinase recognition sequences (RRSs) specific to site specific
CC standard DNA segment. One of the DNA segment (disruption cassette) is
CC inserted in antisense orientation relative to the transcriptional
CC orientation of the gene to be trapped. The other DNA segment (selection
CC cassette) is inserted in sense direction relative to the transcriptional
CC orientation of the gene to be trapped. The cell comprising the gene
CC trapping construct is useful for the identification and/or isolation of
CC genes. The transgenic organism comprising the gene trapping construct is
CC useful to study gene function at various developmental stages. The gene
CC trapping construct is useful for mutationally inactivating all cellular
CC genes. The present sequence is Escherichia coli beta-galactosidase
CC reporter gene of disruption cassette functional DNA segment of gene
CC trapping construct.

XX Sequence 3652 BP; 830 A; 971 C; 1000 G; 851 T; 0 other;

Query Match 21.4%; Score 146; DB 22; Length 3652;
Best Local Similarity 100.0%; Pred. No. 6.1e-33;
Matches 146; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 537 gaattccgcgatactagcgggtccaggagtcgtcgccaccaatccccatatggaaacc 596
DB 3207 GAATTCGCCGATACTAGCGGGTCCAGGAGTCGTCCGCCACCAATCCCCATATGGAACC 3148

QY 597 gtcgatattcagccatgtgcctttcttcgcgcgagcagatggcgatggctgtttccat 656
DB 3147 GTCGATATTACGCCATGTCCTTCTCCGCGTCCAGCAGATGGCGATGGCTGGTTTCCAT 3088

QY 657 cagttgctgttgactgtagcgctga 682
DB 3087 CAGTTGCTGTGACTGTAGCGGCTGA 3062

RESULT 10
AAF85070/c
ID AAF85070 standard; DNA; 3652 BP.

XX
XX
AC AAF85070;

XX
DT 09-JUL-2001 (first entry)

XX Nucleotide sequence of a DNA fragment from Escherichia coli.

DE
XX
XX Gene trapping construct; conditional mutation; recombinase; ss.

KW
XX
XX Escherichia coli.

OS
XX
XX EP1092768-A1.

XX
XX
XX 18-APR-2001.

XX 16-OCT-1999; 99EP-0120592.

PR 16-OCT-1999; 99EP-0120592.

XX
XX (ARTE-) ARTEMIS PHARM GMBH.

PA (FRAN-) FRANKEN BIOTECHNOLOGIE AG.

XX
XX Kuehn R, Von Melchner H;

XX
XX WPI; 2001-275205/29.

XX New gene trapping construct, useful for mutationally inactivating all
PT genes, comprises functional DNA segment inserted in antisense direction
PT relative to gene to be trapped -
XX
XX Disclosure; Page 11-12; 20pp; English.

XX The specification describes a gene trapping construct which is
CC capable of causing conditional mutations in genes. The construct
CC comprises a functional DNA segment inserted in antisense direction
CC relative to the gene to be trapped, where the functional DNA segment
CC is flanked by two recombinase recognition sequences which are specific
CC to a site specific recombinase capable of inverting a double standard
CC DNA segment. The gene trapping construct is useful for identification
CC and isolation of genes. It may be used in methods to study gene function
CC at various developmental stages. The construct is also useful for
CC mutationally inactivating all cellular genes. The present sequence
CC represents a DNA fragment from Escherichia coli, which is used to
CC produce constructs of the invention.

XX Sequence 3652 BP; 830 A; 971 C; 1000 G; 851 T; 0 other;

Query Match 21.4%; Score 146; DB 22; Length 3652;
Best Local Similarity 100.0%; Pred. No. 6.1e-33;
Matches 146; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 537 gaattccgcgatactagcgggtccaggagtcgtcgccaccaatccccatatggaaacc 596
DB 3207 GAATTCGCCGATACTAGCGGGTCCAGGAGTCGTCCGCCACCAATCCCCATATGGAACC 3148

QY 597 gtcgatattcagccatgtgcctttcttcgcgcgagcagatggcgatggctgtttccat 656
DB 3147 GTCGATATTACGCCATGTCCTTCTCCGCGTCCAGCAGATGGCGATGGCTGGTTTCCAT 3088

QY 657 cagttgctgttgactgtagcgctga 682

```
Db 3087 CAGTTGCTGTTGACTGTAGCGGCTGA 3062
|||||
RESULT 11
AAx87978/c
ID AAX87978 standard; DNA; 3756 BP.
XX
AC AAX87978;
XX
DT 06-DEC-1999 (first entry)
XX
DE HCMV UL45 promoter:lacZ gene fusion.
XX
KW HCMV: UL45 gene; promoter; lacZ gene; infection; diagnosis;
herpesvirus; ss.
XX
OS Chimeric - Human cytomegalovirus.
OS Chimeric - Escherichia coli.
XX
FH Key Location/Qualifiers
FT promoter 1..483
FT /*tag= a
FT /*note= "UL45 gene promoter"
FT CDS 484..3714
FT /*tag= b
XX
PN US9598676-A.
XX
PD 28-SEP-1999.
XX
PF 25-APR-1997; 97US-0846026.
XX
PR 18-JUN-1992; 92US-0900279.
PR 28-FEB-1995; 95US-0395673.
XX
PA (UNIW ) UNIV WASHINGTON.
XX
PI Olivo PD;
XX
WP: 1999-561024/47.
P-PSDB; AAY31768.
Genetically engineered cell lines for detecting infectious Herpesvirus
-
Example 17; Fig 13; 44pp; English.
XX
This is the nucleotide sequence of a UL45 promoter/reporter gene
construct composed of the human cytomegalovirus (HCMV) UL45 gene
promoter and 177 bp of the UL45 open reading frame fused to the
Escherichia coli LacZ gene in pCMVUL45lacZ. The predicted product
(see AAY31768) of the chimeric gene is a fusion protein containing 59
amino acids of the UL45 protein fused to the N-terminus of
beta-galactosidase. A claimed in vitro diagnostic method for
detecting infectious HCMV in a specimen comprises: (a) providing a
genetically engineered cell line (especially a mink lung cell line)
stably transformed with DNA containing a beta-gene promoter
sequence of HCMV (especially the UL45 promoter) linked to a
reporter gene (especially a beta-galactosidase gene), whose
expression is dependent upon the presence of HCMV; (b) inoculating
the cell line with a specimen suspected of containing HCMV; (c)
allowing the infectious cycle of the virus to proceed; and (d)
detecting the infectious virus, especially by detecting
beta-galactosidase activity on a fluorogenic substrate or by
histochemical/light microscopy.
XX
Sequence 3756 BP; 836 A; 1033 C; 1079 G; 808 T; 0 other;
XX
Query Match 21.4%; Score 146; DB 20; Length 3756;
Best Local Similarity 100.0%; Pred. No. 6.2e-33;
Matches 146; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 537 gaattccgccgatactgacggggtccaggagtcgtccgcccaatccccatatggaacc 596
|||||
Db 3663 GAATTCCGCCGATAGTACGGGCTCCAGGAGTCGTCGCCCAATCCCATATGGAACC 3604
|||||
Qy 597 gtcgatattcagccatgtgcctttctccgcgtgcagcagatggcgatggctgttccat 656
|||||
Db 3603 GTCGATATTTCAGCCATGTGCTTCTCCGCGTGCAGCAGATGGCGTGGTGGTTCCAT 3544
|||||
Qy 657 cagttgctgtgactgtacgcgtga 682
|||||
Db 3543 CAGTTGCTGTTGACTGTAGCGGCTGA 3518
|||||
RESULT 12
AAX06800/c
ID AAX06800 standard; DNA; 4279 BP.
XX
AC AAX06800;
XX
DT 26-APR-1999 (first entry)
XX
DE Multiple cloning site/Kozak sequence/LacZ gene construct.
XX
KW Gene activation; gene expression; cancer; antisense; therapy;
beta-galactosidase; luciferase; ss.
XX
OS Chimeric - Escherichia coli.
OS Chimeric - synthetic.
XX
FH Key Location/Qualifiers
FT misc_feature 1..64
FT /*tag= a
FT /*function= multiple cloning site
FT misc_signal 65..79
FT /*tag= b
FT /*function= Kozak sequence consensus
FT prim_transcript 80..4279
FT /*tag= c
FT /*standard_name= beta-galactosidase
XX
PN WO9858944-A1.
XX
PD 30-DEC-1998.
XX
PF 24-JUN-1998; 98WO-US13093.
XX
PR 25-JUN-1997; 97US-0050772.
XX
PI (BLAC/) BLACK C A.
XX
PI Black CA;
XX
WP: 1999-095360/08.
PT New masked targetted expression cassette - useful for regulating
gene expression, and preventing neoplastic cell growth
XX
PS Disclosure; Page 22-24; 43pp; English.
XX
This is the multiple cloning site-Kozak-lacZ gene sequence of
armed sense strand plasmid pCI-Neo, which was designed for the
production of the sense RNA strand of a targetted expression
cassette. Alternative flanking sequences (see AAX06801-07)
corresponding to portions of the firefly luciferase mRNA are
inserted into the multiple cloning sites such that transcription
from the T7 promoter yields RNA comprising (from the 5' end):
luciferase segment-Kozak-beta-galactosidase. Sense strand RNA of
a masked targetted cassette is produced by in vitro transcription
of the construct. Antisense sequences (see AAX06808-14)
corresponding to portions of the target molecule are hybridised to
complementary flanking sequences of the sense strand of the
targetted cassette. The hybridised mixture is introduced to an in
```

```

CC  vitro translation mixture containing ribosomes and full-length
CC  firefly luciferase RNA (see AAX06815). After completion of
CC  translation, the mixture is assayed for beta-galactosidase and
CC  luciferase activities. Negative luciferase and positive
CC  beta-galactosidase activity indicates successful inhibition of the
CC  target molecule and successful expression of the gene of interest.
CC  This is an example of a new method for activating genes of
CC  interest that utilises a masked targetted expression cassette
CC  which expresses a gene product only in the presence of a target
CC  molecule. The cassettes are useful for the treatment of disease
CC  and for preventing the proliferation of neoplastic cells. The
CC  cassette only releases the antisense strand and expresses a gene
CC  product in the presence of a target molecule, so is more potent,
CC  reliable, and specific than prior art antisense technology.
XX
SQ  Sequence 4279 BP; 982 A; 1155 C; 1151 G; 991 T; 0 other;

      Query Match      21.4%; Score 146; DB 20; Length 4279;
      Best Local Similarity 100.0%; Pred. No. 6.5e-33;
      Matches 146; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  537 gaattccgcgatactgcagggtccaggaagtgcgccaccacatcccatatggaacc 596
      |||||||
Db  3785 GAATTCGCGGATAGTACGGGGTCCAGAGTGTGTCGCCACCAATCCCATATGGAAACC 3726

QY  597 gtogatattaccgcatgtgccttcttcgcggtgcagcagatggcgatggctgtttccat 656
      |||||||
Db  3725 GTCGATATTACGCCCATGTGCCTTCTTCGCGTGCAGCAGATGCGATGGCTGGTTTCCAT 3666

QY  657 cagttgctttgaactgtagcggctga 682
      |||||||
Db  3665 CAGTTGCTGTTGACTGTAGCGGCTGA 3640

RESULT 13
AAA09085/C
ID AAA09085 standard; DNA; 4487 BP.
XX
AC AAA09085;
XX
XX
DT 10-AUG-2000 (first entry)
XX
DE AdRSV-beta-galactosidase vector region A.
XX
KW Region A; replication-deficient; vector; lacZ; beta-galactosidase;
KW Rous Sarcoma Virus; RSV; promoter; cytotoxicity; cytostatic; pro-drug;
KW prostate cancer; gene therapy; ss.
XX
OS Chimeric - Adenovirus type 5.
OS Chimeric - Rous Sarcoma Virus.
OS Chimeric - Escherichia coli.
XX
PN WO200020038-A1.
XX
PD 13-APR-2000.
XX
PF 01-OCT-1999; 99WO-US20907.
XX
PR 02-OCT-1998; 98US-0165730.
XX
PA (GENO-) GENOTHERAPEUTICS INC.
XX
PI Steiner MS;
XX
XX
DR WPI; 2000-303646/26.
XX
PT Inducing cellular cytotoxicity of tumor cell comprises introducing
PT replication-deficient adenovirus type 5 expression vector containing
PT gene encoding for enzyme having ability to convert nontoxic prodrug
PT into cancer killing drug
XX
PS Claim 4; Fig 18B; 178pp; English.

```

PT into cancer killing drug
 XX Claim 4; Fig 19B; 178pp; English.
 PS This sequence comprises Region A of a replication-deficient adenovirus
 XX type 5 vector containing a lacZ gene (encoding beta-galactosidase
 CC (beta-gal)) under the control of the rat probasin promoter.
 CC Inducing cellular cytotoxicity of a tumor cell comprises introducing a
 CC replication-deficient adenovirus type 5 expression vector comprising a
 CC gene that encodes for an enzyme that has the ability to convert a
 CC non-toxic pro-drug into a cancer killing drug which then destroys cancer
 CC cells. The adenovirus genome preferably has a deletion in an E1 and E3
 CC region and an insertion within the region of a nucleic acid encoding
 CC Escherichia coli beta-gal under the control of a
 CC promoter. The pro-drug active site is masked by beta-gal. Functional
 CC beta-gal is expressed from the vector so as to activate the pro-drug into
 CC an agent toxic to the cells. Beta-gal can be under the control of a
 CC Rous Sarcoma Virus (RSV), probasin (PB), Prostate Specific Antigen (PSA)
 CC or Mouse Mammary Tumor Virus (MMTV) promoter. The vectors provide a
 CC novel way to treat prostate cancer by gene therapy.
 XX Sequence 5201 BP; 1318 A; 1248 C; 1345 G; 1290 T; 0 other;
 SQ

Query Match 21.4%; Score 146; DB 21; Length 5201;
 Best Local Similarity 100.0%; Pred. No. 7e-33;
 Matches 146; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 537 gaattccgccgatactgacgggctccaggagtcgtccgccccaatccccatggaacc 596
 Db 4057 GAATTCCGCCGATACTGACGGGCTCCAGGAGTCGTCCGCCACCAATCCCATATGGAAACC 3998

Qy 597 gtcatattcagccatgtgccttctccgcgtgcagcagatggcgatggctgtttccat 656
 Db 3997 GTCGATATTCAGCCATGTGCCTTCTTCCGCGTCAGCAGATGGCGATGGCTGTGTTCCAT 3938

Qy 657 cagttgctgttgactgtagcggtga 682
 Db 3937 CAGTTGCTGTTGACTGTAGCGGCTGA 3912

RESULT 15
 AAA09089/c
 ID AAA09089 standard; DNA; 5346 BP.
 AC AAA09089;
 XX
 DT 10-AUG-2000 (first entry)
 XX
 DE AdPSA-beta-galactosidase vector region A.
 XX
 KW Region A; replication-deficient; vector; lacZ; beta-galactosidase; PSA;
 KW prostate specific antigen; promoter; cytotoxicity; cytostatic; pro-drug;
 KW prostate cancer; gene therapy; ss.
 XX
 OS Chimeric - Adenovirus type 5.
 OS Chimeric - Homo sapiens.
 OS Chimeric - Escherichia coli.
 XX
 PN W0200020038-A1.
 XX
 PD 13-APR-2000.
 XX
 PF 01-OCT-1999; 99WO-US20907.
 XX
 PR 02-OCT-1998; 98US-0165730.
 XX
 PA (GENO-) GENOTHERAPEUTICS INC.
 XX
 PI Steiner MS;
 XX
 DR WPI; 2000-303646/26.
 XX

PT Inducing cellular cytotoxicity of tumor cell comprises introducing
 PT replication-deficient adenovirus type 5 expression vector containing
 PT gene encoding for enzyme having ability to convert nontoxic prodrug
 PT into cancer killing drug
 XX Claim 4; Fig 20B; 178pp; English.
 PS This sequence comprises Region A of a replication-deficient adenovirus
 XX type 5 vector containing a lacZ gene (encoding beta-galactosidase
 CC (beta-gal)) under the control of the prostate specific antigen (PSA)
 CC promoter.
 CC Inducing cellular cytotoxicity of a tumor cell comprises introducing a
 CC replication-deficient adenovirus type 5 expression vector comprising a
 CC gene that encodes for an enzyme that has the ability to convert a
 CC non-toxic pro-drug into a cancer killing drug which then destroys cancer
 CC cells. The adenovirus genome preferably has a deletion in an E1 and E3
 CC region and an insertion within the region of a nucleic acid encoding
 CC Escherichia coli beta-gal under the control of a
 CC promoter. The pro-drug active site is masked by beta-gal. Functional
 CC beta-gal is expressed from the vector so as to activate the pro-drug into
 CC an agent toxic to the cells. Beta-gal can be under the control of a
 CC Rous Sarcoma Virus (RSV), probasin (PB), PSA or Mouse Mammary Tumor Virus
 CC (MMTV) promoter. The vectors provide a novel way to treat prostate
 CC cancer by gene therapy.
 XX Sequence 5346 BP; 1295 A; 1283 C; 1447 G; 1321 T; 0 other;
 SQ

Query Match 21.4%; Score 146; DB 21; Length 5346;
 Best Local Similarity 100.0%; Pred. No. 7.1e-33;
 Matches 146; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 537 gaattccgccgatactgacgggctccaggagtcgtccgccccaatccccatggaacc 596
 Db 4204 GAATTCCGCCGATACTGACGGGCTCCAGGAGTCGTCCGCCACCAATCCCATATGGAACC 4145

Qy 597 gtcatattcagccatgtgccttctccgcgtgcagcagatggcgatggctgtttccat 656
 Db 4144 GTCGATATTCAGCCATGTGCCTTCTTCCGCGTCAGCAGATGGCGATGGCTGTGTTCCAT 4085

Qy 657 cagttgctgttgactgtagcggtga 682
 Db 4084 CAGTTGCTGTTGACTGTAGCGGCTGA 4059

Search completed: December 27, 2001, 13:14:57
 Job time: 14473 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 27, 2001, 13:06:53 ; Search time 178.26 Seconds
(without alignments)
866.476 Million cell updates/sec

Title: US-09-715-036-3

Perfect score: 682

Sequence: 1 gacacagacacactgtaat.....ctgttgacttagcgctga 682

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 11323899 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_NA.*

1: /cgn2_6/ptodata/2/ina/5A.COMB.seq.*

2: /cgn2_6/ptodata/2/ina/5B.COMB.seq.*

3: /cgn2_6/ptodata/2/ina/6A.COMB.seq.*

4: /cgn2_6/ptodata/2/ina/6B.COMB.seq.*

5: /cgn2_6/ptodata/2/ina/PCTUS.COMB.seq.*

6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	146	21.4	3152	1	US-07-924-028A-3
C 2	146	21.4	3365	1	US-07-789-915A-1
C 3	146	21.4	3365	1	US-08-005-002C-1
C 4	146	21.4	3365	1	US-08-005-002C-1
C 5	146	21.4	8710	1	US-08-487-203A-1
C 6	146	21.4	8710	1	US-08-480-882B-3
C 7	146	21.4	8710	1	US-08-480-210-3
C 8	146	21.4	9019	1	US-08-480-882B-4
C 9	146	21.4	9019	1	US-08-480-210-4
C 10	146	21.4	9641	2	US-08-374-483-3
C 11	146	21.4	10241	2	US-08-508-448C-16
C 12	146	21.4	10281	2	US-08-816-155B-1
C 13	146	21.4	10281	2	US-08-079-587-1
C 14	146	21.4	34382	2	US-08-374-483-6
C 15	145	21.3	7852	3	US-08-836-022A-2
C 16	145	21.3	7852	4	US-09-427-048A-2
C 17	145	21.3	8313	1	US-08-232-463-2
C 18	145	21.3	8509	1	US-08-462-014-1
C 19	145	21.3	8509	4	US-08-973-334-4
C 20	145	21.3	8775	1	US-09-563-869A-4
C 21	145	21.3	9454	1	US-08-232-463-5
C 22	145	21.3	9454	1	US-08-232-463-3
C 23	145	21.3	9661	3	US-08-232-463-4
C 24	145	21.3	9661	3	US-08-716-351A-3
C 25	145	21.3	9890	1	US-08-232-463-18
C 26	145	21.3	9916	1	US-08-232-463-17
C 27	145	21.3	9917	3	US-08-232-463-16
C 28	145	21.3	10306	3	US-08-716-351A-4

C 28	145	21.3	10398	2	US-08-331-384-1	Sequence 1, Appli
C 29	145	21.3	10398	2	US-08-708-188-1	Sequence 1, Appli
C 30	145	21.3	10398	2	US-08-836-087-1	Sequence 1, Appli
C 31	145	21.3	10398	4	US-09-246-320-1	Sequence 1, Appli
C 32	145	21.3	10398	4	US-09-242-743-1	Sequence 1, Appli
C 33	145	21.3	10408	1	US-08-232-463-6	Sequence 6, Appli
C 34	145	21.3	10408	1	US-08-232-463-7	Sequence 7, Appli
C 35	145	21.3	10970	3	US-08-716-351A-5	Sequence 5, Appli
C 36	145	21.3	35408	4	US-08-973-334-3	Sequence 3, Appli
C 37	145	21.3	35408	4	US-09-563-869A-3	Sequence 3, Appli
C 38	145	21.3	35408	4	US-08-549-489-3	Sequence 3, Appli
C 39	143.4	21.0	7897	3	US-08-836-022A-1	Sequence 1, Appli
C 40	143.4	21.0	7897	4	US-09-427-048A-1	Sequence 1, Appli
C 41	142.8	20.9	4810	3	US-08-852-629-11	Sequence 11, Appl
C 42	142.8	20.9	4838	3	US-08-852-629-15	Sequence 15, Appl
C 43	142.8	20.9	19182	2	US-08-850-880-11	Sequence 11, Appl
C 44	142.8	20.9	19182	2	US-08-944-916-11	Sequence 11, Appl
C 45	140.2	20.6	619	5	PCT-US95-13658-5	Sequence 5, Appli

ALIGNMENTS

RESULT 1
US-07-924-028A-3/C
; Sequence 3, Application US/07924028A
; Patent No. 5470573
; GENERAL INFORMATION:
; APPLICANT: Lubitz Werner, Szostak, Michael P.
; TITLE OF INVENTION: CARRIER-BOUND RECOMBINANT PROTEINS, PROCESS
; TITLE OF INVENTION: FOR THE PRODUCTION AND USE AS IMMUNOGENS AND VACCINES
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felfe & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; COUNTRY: USA
; ZIP: 10022

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
COMPUTER: IBM PS/2
OPERATING SYSTEM: PC-DOS
SOFTWARE: Wordperfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/924,028A
FILING DATE: 30-SEP-1992
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/EP91/00308
FILING DATE: 02-FEB-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: DE 40 05 874
FILING DATE: 24-FEB-1990
ATTORNEY/AGENT INFORMATION:
NAME: Hanson, No. 5470573man D.
REGISTRATION NUMBER: 30,946
REFERENCE/DOCKET NUMBER: HUBR 1027
TELEPHONE: (212) 688-9200
TELEFAX: (212) 838-3884
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 3152 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-07-924-028A-3

Query Match 21.4%; Score 146; DB 1; Length 3152;
Best Local Similarity 100.0%; Pred. No. 1.6e-33;
Matches 146; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	537	gaattccgcgcatactagcgggtctccaggagtgcgtccaccaatccccatatgaaacc	596
Db	3036	GAATTCCGCGGATCTAGCGGGTCCAGGAGTCGTGCAGCAATAATGCCCATATGSAACC	2977
QY	597	gtcgatatccagccatgtaccttcttcgcgtgcagcagatggcgatgctggtttccat	656
Db	2976	GTCGATATTCAGCCATGTCCCTCTCTCCGCTGCACAGATGGCGATGGCTGGTTCCAT	2917
QY	657	cagtgctgttgactgtacggctga	682
Db	2916	CAGTGCTGTTGACTGTAGCGGCTGA	2891

```

RESULT      2
US-07-789-915A-1/c
; Sequence 1, Application US/07789915A
; Patent No. 5212058
; GENERAL INFORMATION:
; APPLICANT: Baker, Rohan T.
; APPLICANT: Tobias, John W.
; APPLICANT: Varshavsky, Alexander
; TITLE OF INVENTION: Ubiquitin-Specific Proteases
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02173

```

02175
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/07/789,915A
 FILING DATE: 19911108
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Brook, David E.
 REGISTRATION NUMBER: 22,592
 REFERENCE/DOCKET NUMBER: MIT-5091AA
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 617-861-6240
 TELEFAX: 617-861-9540

```

, LIBRARY: PDB ID: 1J36
, INFORMATION FOR SEQ ID NO: 1:
,-----
, SEQUENCE CHARACTERISTICS:
, LENGTH: 3365 base pairs
, TYPE: NUCLEIC ACID
, STRANDEDNESS: double
, TOPOLOGY: linear
,-----
, FEATURE:
, NAME/KEY: CDS
, LOCATION: 1..3363
US-07-789-915A-1

```

```
Query Match      21.4%; Score 146; DB 1; Length 3365;
Best Local Similarity 100.0%; Pred. No. 1.6e-33;
Matches 146: Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

Qy	537	gaattccgcgatactactgacggctcagggagctgctgcacccaatccccatgatgaaacc	596
Db	3315	GAATTCGGCGATACTAGGGGCTCAGGAGTCGTCCGCACCAATCCCATATGGAAACC	3256
Qy	597	gtcgatatccagccatgctcctcttcacgcgcagcagatggcgatgctggtttccat	656
Db	3255	GTGATATTACGCCATGTGCTTCTTCGCGTCACGAGATGGCGATGGCTGGTTCCAT	3196
Qy	657	cagttgctgttgactgacggccta	682

```

Db      3195 CAGTTCCTGTTGACTGTACCGCCTGA 3170

RESULT          3
US-08-005-002C-1/c
; Sequence 1, Application US/08005002C
; Patent No. 5494818
; GENERAL INFORMATION:
; APPLICANT: Baker, Rohan T.
; APPLICANT: Tobias, John W.
; APPLICANT: Varshavsky, Alexander
; TITLE OF INVENTION: Ubiquitin-Specific Proteases
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kevin M. Farrell, P.C.
; STREET: P.O. Box 999
; CITY: York Harbor
; STATE: Maine
; COUNTRY: U.S.A.
; ZIP: 03911
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/005,002C
; FILING DATE: 15-JAN-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/789,915
; FILING DATE: 08-NOV-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Farrell, Kevin M.
; REGISTRATION NUMBER: 35,505
; REFERENCE/DOCKET NUMBER: MIT-5091AAZ
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 207-363-0558
; TELEFAX: 207-363-0528
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3365 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..3363
; PS-08-005-002C-1

```

	Query Match	21.4%	Score 146;	DB 1;	Length 3365;
	Best Local Similarity	100.0%;	Pred. No. 1.6e-33;		
	Matches 146;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	537	gaattccgcgcatactgacgggtctcaggagtcctcgccaccatcccccatatggaacc	596		
Dd	3315	GAATTCGCCGCATACTGACGGGCTCCAGGAGTGTGCGCACCAATCCCACATATGGAACC	3256		
Qy	597	gtcgatatccagccatgtgccctttctccgctgcagcagatggcgatggctgggttccat	656		
Dd	3255	GTCGATATTACCAATGTGCCTTTCTCCGCTGTCAGCAGATGGCGATGGTGCTGGTTCCAT	3196		
Qy	657	cagttgtcgttgactgtagcggtga	682		
Dd	3195	CAGTTGCTGTTTGA CTGTAGCGGCTGA	3170		

RESULT 4
US-08-487-203A-1/c
; Sequence 1, Application US/08487203A
; Patent No. 5683904
; GENERAL INFORMATION:

APPLICANT: Baker, Rohan T.
APPLICANT: Tobias, John W.
APPLICANT: Varshavsky, Alexander
TITLE OF INVENTION: Ubiquitin-Specific Proteases
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Kevin M. Farrell, P.C.
STREET: P.O. Box 999
CITY: York Harbor
STATE: Maine
COUNTRY: U.S.A.
ZIP: 03911
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/487,203A
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/005,002
FILING DATE: 15-JAN-1993
ATTORNEY/AGENT INFORMATION:
NAME: Farrell, Kevin M.
REGISTRATION NUMBER: 35,505
REFERENCE/DOCKET NUMBER: MIT-5091A3Z
TELECOMMUNICATION INFORMATION:
TELEPHONE: 207-363-0558
TELEFAX: 207-363-0528
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3365 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 1..3363
US-08-487-203A-1

Query Match 21.4%; Score 146; DB 1; Length 3365;
Best Local Similarity 100.0%; Pred. No. 1.6e-33;
Matches 146; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 537 gaattccgcgatactgacgggtccaggagtcgtcgccaccacatccccatattggaacc 596
|||||
Db 3315 GAATTCGCCGATACTGACGGGCTCCAGGAGTCGTCCGCCACCAATCCCCATATGGAACC 3256
QY 597 gtccattaccagcatgtgctcttcctccgctgacagcagatggcgatggctgttccat 656
|||||
Db 3255 GTCGATATTCAGCATGTGCTCTTCCTCCGCTGCAGCAGATGCGATGCGTGGTTTCCAT 3196
QY 657 cagttgctgtgactgtagcggtga 682
|||||
Db 3195 CAGTTGCTGTGACTGTAGCGGCTGA 3170

RESULT 5
US-08-480-882B-3
Sequence 3, Application US/08480882B
Patent No. 5656275
GENERAL INFORMATION:
APPLICANT: WASMOEN, TERRI
APPLICANT: CHAVEZ, LLOYD
APPLICANT: CHU, HSIEH-JUE
TITLE OF INVENTION: RECOMBINANT RACCOON POX VIRUSES
TITLE OF INVENTION: AND THEIR USE AS AN EFFECTIVE VACCINE AGAINST FELINE
TITLE OF INVENTION: INFECTIOUS PERITONITIS VIRUS DISEASE
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:

ADDRESSEE: Darby & Darby, P.C.
STREET: 805 Third Avenue 27th Floor
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/480,882B
FILING DATE: 07-JUN-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/125,516
FILING DATE: 22-SEP-1993
ATTORNEY/AGENT INFORMATION:
NAME: Green, Reza
REGISTRATION NUMBER: 38,475
REFERENCE/DOCKET NUMBER: 0632/08669
TELECOMMUNICATION INFORMATION:
TELEPHONE: 2125277700
TELEFAX: 2127536237
TELEX: 236687
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 8710 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Feline infectious peritonitis virus
IMMEDIATE SOURCE:
CLONE: pscilf1
US-08-480-882B-3

Query Match 21.4%; Score 146; DB 1; Length 8710;
Best Local Similarity 100.0%; Pred. No. 2.4e-33;
Matches 146; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 537 gaattccgcgatactgacgggtccaggagtcgtcgccaccacatccccatattggaacc 596
|||||
Db 3102 GAATTCGCCGATACTGACGGGCTCCAGGAGTCGTCCGCCACCAATCCCCATATGGAACC 3161
QY 597 gtccattaccagcatgtgctcttcctccgctgacagcagatggcgatggctgttccat 656
|||||
Db 3162 GTCGATATTCAGCATGTGCTCTTCCTCCGCTGCAGCAGATGCGATGCGTGGTTTCCAT 3221
QY 657 cagttgctgtgactgtagcggtga 682
|||||
Db 3222 CAGTTGCTGTGACTGTAGCGGCTGA 3247

RESULT 6
US-08-480-210-3
Sequence 3, Application US/08480210
Patent No. 5770211
GENERAL INFORMATION:
APPLICANT: WASMOEN, TERRI
APPLICANT: CHAVEZ, LLOYD
APPLICANT: CHU, HSIEH-JUE
TITLE OF INVENTION: RECOMBINANT RACCOON POX VIRUSES
TITLE OF INVENTION: AND THEIR USE AS AN EFFECTIVE VACCINE AGAINST FELINE
TITLE OF INVENTION: INFECTIOUS PERITONITIS VIRUS DISEASE
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Darby & Darby, P.C.
STREET: 805 Third Avenue 27th Floor
CITY: New York

STATE: NY
COUNTRY: USA
ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSEQ Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/480,210
FILING DATE: 07-JUN-1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/125,516
FILING DATE: 22-SEP-1993
ATTORNEY/AGENT INFORMATION:
NAME: Green, Reza
REGISTRATION NUMBER: 38,475
REFERENCE/DOCKET NUMBER: 0632/18669-US1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 2125277700
TELEFAX: 2127536237
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 8710 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Feline infectious peritonitis virus
IMMEDIATE SOURCE:
CLONE: psc11f
US-08-480-210-3

Query Match 21.4%; Score 146; DB 1; Length 8710;
Best Local Similarity 100.0%; Pred. No. 2.4e-33; Indels 0; Gaps 0;
Matches 146; Conservative 0; Mismatches 0;
QY 537 gaattccgcgactgacgggtccagagtgctgcgccaccaatcccccatatggaacc 596
|||||
Db 3102 GAATTCGGCGGATCTGACGGGCTCCAGAGTGTGCGCCACCAATCCCATATGGAACC 3161
|||||
QY 597 gtcgatattcagccatgtgccttcttcgcggtgcagcagatggcgatggctggttccat 656
|||||
Db 3162 GTCGATATTCAGCCATGTGCTTCTTCCGCGTGCAGCAGATGGCGATGGCTGTTCCAT 3221
|||||
QY 657 cagttgctgtgactgtagcgctga 682
|||||
Db 3222 CAGTTGCTGTGACTGTAGCGGCTGA 3247
|||||

RESULT 7
US-08-480-882B-4
; Sequence 4, Application US/08480882B
; Patent No. 5656275
; GENERAL INFORMATION:
; APPLICANT: WASMOEN, TERRI
; APPLICANT: CHAVEZ, LLOYD
; APPLICANT: CHU, HSIEH-JUE
; TITLE OF INVENTION: RECOMBINANT RACCOON POX VIRUSES
; TITLE OF INVENTION: AND THEIR USE AS AN EFFECTIVE VACCINE AGAINST FELINE
; TITLE OF INVENTION: INFECTIOUS PERITONITIS VIRUS DISEASE
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Darby & Darby, P.C.
; STREET: 805 Third Avenue
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10022

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSEQ Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/480,882B
FILING DATE: 07-JUN-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/125,516
FILING DATE: 22-SEP-1993
ATTORNEY/AGENT INFORMATION:
NAME: Green, Reza
REGISTRATION NUMBER: 38,475
REFERENCE/DOCKET NUMBER: 0632/08669
TELECOMMUNICATION INFORMATION:
TELEPHONE: 2125277700
TELEFAX: 2127536237
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 9019 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Feline immunodeficiency virus
IMMEDIATE SOURCE:
CLONE: psc11e1
US-08-480-882B-4

Query Match 21.4%; Score 146; DB 1; Length 9019;
Best Local Similarity 100.0%; Pred. No. 2.4e-33; Indels 0; Gaps 0;
Matches 146; Conservative 0; Mismatches 0;
QY 537 gaattccgcgactgacgggtccagagtgctgcgccaccaatcccccatatggaacc 596
|||||
Db 3101 GAATTCGGCGGATCTGACGGGCTCCAGAGTGTGCGCCACCAATCCCATATGGAACC 3160
|||||
QY 597 gtcgatattcagccatgtgccttcttcgcggtgcagcagatggcgatggctggttccat 656
|||||
Db 3161 GTCGATATTCAGCCATGTGCTTCTTCCGCGTGCAGCAGATGGCGATGGCTGTTCCAT 3220
|||||
QY 657 cagttgctgtgactgtagcgctga 682
|||||
Db 3221 CAGTTGCTGTGACTGTAGCGGCTGA 3246
|||||

RESULT 8
US-08-480-210-4
; Sequence 4, Application US/08480210
; Patent No. 5770211
; GENERAL INFORMATION:
; APPLICANT: WASMOEN, TERRI
; APPLICANT: CHAVEZ, LLOYD
; APPLICANT: CHU, HSIEH-JUE
; TITLE OF INVENTION: RECOMBINANT RACCOON POX VIRUSES
; TITLE OF INVENTION: AND THEIR USE AS AN EFFECTIVE VACCINE AGAINST FELINE
; TITLE OF INVENTION: INFECTIOUS PERITONITIS VIRUS DISEASE
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Darby & Darby, P.C.
; STREET: 805 Third Avenue
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible

Qy	537	gaattccgcgcgatactaacggggtccaggagtcgtgcgcaccaatcccatatgaaac	596
Db	464	GAATTCGCCGCGATACGACGGGTCGACGGAGTCGCGCCACCAATCCCATATGGAAAC	523
Qy	597	gtcgatatccagccatcgtgccttcttcgcgctcgcgcagatggcgatgctggtttccat	656
Db	524	GTCGATATTCAGGCATGTCGCTTCTTCGCGCTGCACAGATGCGATGGCGGTTTTCAT	583

584 CAGTTGCTGTTGACTGTAGCGGCTGA 609

RESULT 10
US-08-508-448C-16
; Sequence 16, Application US/08500448C
; Patent No. 5804410
; GENERAL INFORMATION:
; APPLICANT: Kazuyoshi YAMAOKA et al.
; TITLE OF INVENTION: NUCLEIC ACID SEQUENCE
; OF INVENTION: TRYPSIN-LIKE ENZYME AN
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack
; STREET: 905 Fifteenth Street, N.W., #700
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.

STATE: D.C. U.S.A.
COUNTRY: U.S.A.
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/508,448C
FILING DATE: July 28, 1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Warren M. Cheek, Jr.
REGISTRATION NUMBER: 33,367
REFERENCE/DOCKET NUMBER:
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-371-8850
TELEFAX:
TELEX:
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:

LENGTH: 10241 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: cyclic
MOLECULE TYPE: other nucleic acid (vector DNA)
US-08-508-448C-16

Query Match
Best Local Similarity 21.4%; Score 146; DB 1; Length 10241;
Matches 146; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 537 gaattccgcgatactgacgggtccaggagtcgtccaccacccatcccatatggaacc 596
|||||
Db 2426 GAATTCGCCGATACTGACGGGCTCCAGGAGTCGTCGCCACCAATCCCAATATGGAACC 2485
Qy 597 gtcgatatccagccatgtgccttcttcctccgcgtgcagcagatggcgatggctgtttccat 656
|||||
Db 2486 GTCGATATTACGCCATGTGCTTCTCCGCGTCGACGAGATGGCGATGGCTGTTCAT 2545
Qy 657 cagttgctgtgactagcggctga 682
|||||
Db 2546 CAGTGTCTGTGACTGAGCGGCTGA 2571

RESULT 11
US-08-816-155B-1
; Sequence 1, Application US/08816155B
; Patent No. 5990091
; GENERAL INFORMATION:
; APPLICANT: TARTAGLIA, JAMES
; APPLICANT: COX, WILLIAM I.
; APPLICANT: GETTIG, RUSSELL R.
; APPLICANT: MARTINEZ, HECTOR
; APPLICANT: PAOLETTI, ENZO
; APPLICANT: PINCUS, STEVEN E.
; TITLE OF INVENTION: VECTORS HAVING ENHANCED EXPRESSION, AND
; NUMBER OF SEQUENCES: 48
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FROMMER LAWRENCE & HAUG LLP
; STREET: 745 FIFTH AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10151

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/816,155B
FILING DATE: 12-MAR-1997
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: KOWALSKI, THOMAS J.
REGISTRATION NUMBER: 32,147
REFERENCE/DOCKET NUMBER: 454310-2990
TELEPHONE: 212-588-0800
TELEFAX: 212-588-0500
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 10281 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-816-155B-1

Query Match 21.4%; Score 146; DB 2; Length 10281;

Best Local Similarity 100.0%; Pred. No. 2.5e-33;
Matches 146; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 537 gaattccgcgatactgacgggtccaggagtcgtccaccacccatcccatatggaacc 596
|||||
Db 7659 GAATTCGCCGATACTGACGGGCTCCAGGAGTCGTCGCCACCAATCCCAATATGGAACC 7718
Qy 597 gtcgatatccagccatgtgccttcttcctccgcgtgcagcagatggcgatggctgtttccat 656
|||||
Db 7719 GTCGATATTACGCCATGTGCTTCTTCGCGTCGACGAGATGGCGATGGCTGTTCAT 7778
Qy 657 cagttgctgtgactagcggctga 682
|||||
Db 7779 CAGTGTCTGTGACTGAGCGGCTGA 7804

RESULT 12
US-09-079-587-1
; Sequence 1, Application US/09079587
; Patent No. 6130066
; GENERAL INFORMATION:
; APPLICANT: TARTAGLIA, JAMES
; APPLICANT: COX, WILLIAM I.
; APPLICANT: GETTIG, RUSSELL R.
; APPLICANT: MARTINEZ, HECTOR
; APPLICANT: PAOLETTI, ENZO
; APPLICANT: PINCUS, STEVEN E.
; TITLE OF INVENTION: VECTORS HAVING ENHANCED EXPRESSION, AND
; NUMBER OF SEQUENCES: 48
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FROMMER LAWRENCE & HAUG LLP
; STREET: 745 FIFTH AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10151

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/079,587
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/816,155
FILING DATE: 12-MAR-1997
ATTORNEY/AGENT INFORMATION:
NAME: KOWALSKI, THOMAS J.
REGISTRATION NUMBER: 32,147
REFERENCE/DOCKET NUMBER: 454310-2990
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-588-0800
TELEFAX: 212-588-0500
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 10281 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-09-079-587-1

Query Match 21.4%; Score 146; DB 3; Length 10281;
Best Local Similarity 100.0%; Pred. No. 2.5e-33;
Matches 146; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 537 gaattccgcgatactgacgggtccaggagtcgtccaccacccatcccatatggaacc 596
|||||
Db 7659 GAATTCGCCGATACTGACGGGCTCCAGGAGTCGTCGCCACCAATCCCAATATGGAACC 7718

QY 597 gtccatattccagccatgtgcttctccgctgcagcagatgagcgatggctgtttccat 656
Db 7719 GTCGATATTACGCATGTGCTTCTCCGCGTGACGAGATGGCGATGGCTGTTTCCAT 7778
QY 657 cagttgctgttgactgtagcggtga 682
Db 7779 CAGTTGCTGTGACTGTAGCGGTGA 7804

RESULT 13
US-08-374-483-6
; Sequence 6, Application US/08374483
; Patent No. 5880102
; GENERAL INFORMATION:
; APPLICANT: GEORGE, SAMUEL E.
; APPLICANT: BLAZING, MICHAEL A.
; TITLE OF INVENTION: ADENOVIRAL VECTOR SYSTEM
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHYE P.C.
; STREET: 1100 NORTH GLEBE ROAD, 8TH FLOOR
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/374,483
; FILING DATE: 17-JAN-1995
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: WILSON, MARY J.
; REGISTRATION NUMBER: 32,955
; REFERENCE/DOCKET NUMBER: 1579-83
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4000
; TELEFAX: (703) 816-4100
; TELEX: 200797 NIXN UR
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 34382 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-374-483-6

Query Match 21.4%; Score 146; DB 2; Length 34382;
Best Local Similarity 100.0%; Pred. No. 4.1e-33;
Matches 146; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 537 gaattccgcgatactgacgggtccagagtgctgcgccaccaatcccatatggaacc 596
Db 384 GAATTCGCGGATACGAGGCTCCAGGAGTGTGCGCCACCAATCCCATATGGAACC 443
QY 597 gtcgatattccagctgtccttctccgctgcagcagatggcgatggctgtttccat 656
Db 444 GTCGATATTACGCATGTGCTTCTCCGCGTGACGAGATGGCGATGGCTGTTTCCAT 503
QY 657 cagttgctgttgactgtagcggtga 682
Db 504 CAGTTGCTGTGACTGTAGCGGTGA 529

RESULT 14
US-08-836-022A-2/c
; Sequence 2, Application US/08836022A

; Patent No. 6001557
; GENERAL INFORMATION:
; APPLICANT: Trustees of the University of Pennsylvania
; APPLICANT: Wilson, James M.
; APPLICANT: Fisher, Krishna J.
; APPLICANT: Chen, Shu-Jen
; APPLICANT: Weitzman, Matthew
; TITLE OF INVENTION: Improved Adenovirus Virus and
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howson and Howson
; STREET: Spring House Corporate Cntr, P O Box 457
; CITY: Spring House
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19477
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/836,022A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/331,381
; FILING DATE: 28-OCT-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Bak, Mary E.
; REGISTRATION NUMBER: 31,215
; REFERENCE/DOCKET NUMBER: GNPVN.008PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-540-9200
; TELEFAX: 215-540-5818
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7852 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: cdna
US-08-836-022A-2

Query Match 21.3%; Score 145; DB 3; Length 7852;
Best Local Similarity 100.0%; Pred. No. 4.5e-33;
Matches 145; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 538 aattccgcgatactgacgggtccagagtgctgcgccaccaatcccatatggaaccg 597
Db 4907 AATTCCGCGGATACGAGGCTCCAGGAGTGTGCGCCACCAATCCCATATGGAACCG 4848
QY 598 tcgatattccagctgtgcttcttccgctgcagcagatggcgatggctgtttccatc 657
Db 4847 TCGATATTACGCATGTGCTTCTCCGCGTGACGAGATGGCGATGGCTGTTTCCATC 4788
QY 658 agttgctgttgactgtagcggtga 682
Db 4787 AGTTGCTGTGACTGTAGCGGTGA 4763

RESULT 15
US-09-427-048A-2/c
; Sequence 2, Application US/09427048A
; Patent No. 6203975
; GENERAL INFORMATION:
; APPLICANT: Trustees of the University of Pennsylvania
; Wilson, James M.
; Fisher, Krishna J.
; Chen, Shu-Jen
; Weitzman, Matthew
; TITLE OF INVENTION: Improved Adenovirus Virus and

Search completed: December 27, 2001, 13:07:22
Job time: 14058 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using SW model

Run on: December 27, 2001, 13:14:57 ; Search time 486.97 Seconds
(without alignments)
3262.262 Million cell updates/sec

Title: US-09-715-036-4
Perfect score: 1853
Sequence: 1 atgctaacgtctcttgagct.....aaaaaaaaaaaaaaaaaaaa 1853

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 930621 seqs, 428662619 residues
Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: N_Geneseq_1101.*
1: /SID2/gcgdata/geneseq/geneseq/NA1980.DAT:*
2: /SID2/gcgdata/geneseq/geneseq/NA1981.DAT:*
3: /SID2/gcgdata/geneseq/geneseq/NA1982.DAT:*
4: /SID2/gcgdata/geneseq/geneseq/NA1983.DAT:*
5: /SID2/gcgdata/geneseq/geneseq/NA1984.DAT:*
6: /SID2/gcgdata/geneseq/geneseq/NA1985.DAT:*
7: /SID2/gcgdata/geneseq/geneseq/NA1986.DAT:*
8: /SID2/gcgdata/geneseq/geneseq/NA1987.DAT:*
9: /SID2/gcgdata/geneseq/geneseq/NA1988.DAT:*
10: /SID2/gcgdata/geneseq/geneseq/NA1989.DAT:*
11: /SID2/gcgdata/geneseq/geneseq/NA1990.DAT:*
12: /SID2/gcgdata/geneseq/geneseq/NA1991.DAT:*
13: /SID2/gcgdata/geneseq/geneseq/NA1992.DAT:*
14: /SID2/gcgdata/geneseq/geneseq/NA1993.DAT:*
15: /SID2/gcgdata/geneseq/geneseq/NA1994.DAT:*
16: /SID2/gcgdata/geneseq/geneseq/NA1995.DAT:*
17: /SID2/gcgdata/geneseq/geneseq/NA1996.DAT:*
18: /SID2/gcgdata/geneseq/geneseq/NA1997.DAT:*
19: /SID2/gcgdata/geneseq/geneseq/NA1998.DAT:*
20: /SID2/gcgdata/geneseq/geneseq/NA1999.DAT:*
21: /SID2/gcgdata/geneseq/geneseq/NA2000.DAT:*
22: /SID2/gcgdata/geneseq/geneseq/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1853	100.0	1855	22	AAF90338 Peanut allergen Ar
2	1391.2	75.1	1524	20	AAZ22280 Ara h 3 allergen e
3	1391.2	75.1	1524	20	AAZ06384 Peanut allergen A
4	1388	74.9	1524	22	AA508540 DNA encoding anaph
5	527.2	28.5	1446	19	AAV17567 Coding sequence fo
6	516	27.8	1743	21	AA292638 CDNA encoding soyb
7	515.6	27.8	1488	19	AAV17565 Coding sequence fo
8	510.6	27.6	1712	11	AAO05359 Glycitin subunit p
9	508.2	27.4	1458	19	AAV17566 Coding sequence fo
10	501.6	27.1	1746	11	AAO05360 Glycitin subunit p
11	201.8	10.9	1551	19	AAV17569 Coding sequence fo

12	201.8	10.9	1786	11	AAO05357 Glycitin subunit p
13	191.4	10.3	1786	7	AAAG0939 Sequence encoding
14	168.2	9.1	1899	11	AAO05358 Glycitin subunit p
15	158.8	8.6	1646	9	AAAB2246 Rice storage prote
16	152.2	8.2	1689	19	AAV17568 Coding sequence fo
17	152.2	8.2	1896	7	AAAG0940 Sequence encoding
18	138.2	7.5	1706	20	AAV37335 Coffee storage pro
19	125.8	6.8	1685	15	AAO54819 oat globin A2b cod
20	121	6.5	1544	15	AAO54818 Modified oat globl
21	118.4	6.4	2844	8	AAV1250 Sequence of Vicia
22	113	6.1	3113	12	AAO13670 Cruciferin A gene.
23	109.8	5.9	4999	22	AAAD03030 Flax legumin-like
24	76.4	4.1	936	22	AAAF58252 Oligonucleotide D1
25	76.4	4.1	936	22	AAAF58254 Oligonucleotide D1
26	76.4	4.1	936	22	AAAF58257 Oligonucleotide D1
27	76.4	4.1	936	22	AAAF58259 Oligonucleotide D2
28	76.4	4.1	936	22	AAAF58262 Oligonucleotide D1
29	76.4	4.1	936	22	AAAF58265 Oligonucleotide D1
30	74.4	4.0	936	22	AAAF58252 Oligonucleotide D1
31	74.4	4.0	936	22	AAAF58254 Oligonucleotide D1
32	74.4	4.0	936	22	AAAF58257 Oligonucleotide D1
33	74.4	4.0	936	22	AAAF58259 Oligonucleotide D2
34	74.4	4.0	936	22	AAAF58262 Oligonucleotide D1
35	74.4	4.0	936	22	AAAF58265 Oligonucleotide D1
36	59	3.2	244	22	AAAF58238 Oligonucleotide D1
37	56.8	3.1	244	22	AAAF58238 Oligonucleotide D1
38	53	2.9	3477	20	AAV37336 Coffee storage pro
39	51.2	2.8	310	22	AAH71505 Human cervical can
40	50.4	2.7	1358	21	AAAC3232 Arabidopsis thalia
41	47	2.5	577	21	AAAC44643 Human cervical can
42	46.2	2.5	655	22	AAH70113 Human colon cancer
43	46	2.5	3179	22	AAH33311 Murine CD39-L4 gen
44	44.8	2.4	6789	22	AAAF63436 Corn CCR4 transcri
45	44.6	2.4	1495	21	AAAZ7985

ALIGNMENTS

RESULT 1	
AAF90338	AAF90338 standard; cDNA; 1855 BP.
XX	AAF90338;
AC	
XX	
DT	23-JUL-2001 (first entry)
XX	
DE	Peanut allergen Ara h4 cDNA.
XX	
KW	Peanut; allergen; Ara h2; Ara h3; transgenic plant; allergy; ss.
XX	
OS	Arachis hypogea.
XX	
FH	Key
FT	CDS 1..1593 Location/Qualifiers
FT	/*tag= a
FT	misc_feature 430..1190
FT	/*tag= b
FT	/note= "PCR amplified fragment"
XX	
PN	WO200136621-A2.
XX	
PD	25-MAY-2001.
XX	
PF	20-NOV-2000; 2000WO-US31657.
XX	
PR	19-NOV-1999; 99US-0167255.
XX	
PA	(UYAL-) UNTIV ALABAMA A & M.
XX	
PI	Dodo HW, Arintzen CJ, Konan KN, Viquez OM;
XX	
DR	WPI; 2001-355630/37.

SD Sequence 1855 BP; 564 A; 430 C; 465 G; 396 T; 0 other;

Matches 1853; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 atgcctaagctcttgaacatttcttgaatttctaattctaacaaactaacacc 60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

18

QY 181 caggagttcgaatgcgcgcgctgcgccctctctcgcttagtccctcgcgcgcaagccct 24

QY 241 cgtatgaccttctactccaatgctccccaagaatcttcattccaacaaagaaagaaatac 30

[illegible]

360

Db 361 cgatatcagttcccaagaccaccaagacgtttgcaagaagaagaccacaagccaacacagcaa 420

Db 421 caagatagtcaccagaaggtgcaccgttccaatgaaggtatctcatcttgaattccacc 480

481
pb

[illegible]

601 aaccacgagcaagagtctctaaggtaccagcaacaagcagacgaagaagc 660

Db 661. ttaccatatagcccatacagcccgcatgctcgccctagacagaaagacgtgaatttcgc 720

[illegible]

yy |aacatctcagcggttcacgcccgggttccctgtaacaagccttcacagtctgacgacaga 840

Db 841 cagattgtgcacaatctgttggtgcgagaaacgagagtgtgaagaagagggagccattgtgacg 900

Db 901 gtgagggaagcctcaagaatcttcagcccaagatggaacgaagacgttcgcgaacgaagaag 960

[illegible]

Db 1021 agagcgggggaatggtatgaagacgacatctgcaccgatgtgtlaaagaacatt 108

QY 1141 gatctaaccttctaactccttagtgccttgacttagtgcctgaatatgqaatctctac 1200

[illegible]

uu 1201 aggaatgcattgttgcctcactacacaccacgcacacagcatcatatgtcatg 1260

Db 1261 aggggacgycgtcaagtggtgacagcaacgycaacacagagtgtacgacgagggag 1320

Dp 1321 ctccaagaggtcacgtctctgtgtgccacagaacttcgacctgtgccttgggaagtccca 1380

1381 agcagaaacttcgaatgacgtatgcattcaagagagacattcaaacgccacagacatgacacacattt 1440

gccccgcgaaccccccacagataaacctcgccggaggaggtcggcttgcgaattcatatgtgc 1500

QY 1501 ctcccaaggagcagcgaagcagcttaagaacaacacccctcaagttcttcgtcca 1560

QY 1561 cctttcagcagctctcgcgaggctgtgtgcttaaaaacgaccagatcttttgcaagcgtg 1620

OY 1621 ttatccactaacataactitttgcacaaatgaataatataatataatgaagaaatatar 1680

1601 1080

Db 1681 agtttaatttagtatgaataagaacaaggcgcatgacgtcttttgtttaaga 1740


```

QY      1450 aactcttcataagataacctccggaggaggtggttccaattcataatggcctcccgaag 1509
        ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||
Db       1381 aactccgcacataagaaccctcggaagagtgtgtccaattcataatggcctccaagg 1440
QY      1510 gacacgaagcagcaacttaagaacaacccccctaagtcttcgttcacatttcag 1569
        ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||
Db       1441 gagcacgaagcagcagcttaagaacaacacccccctaagtcttcgttcacacgtttcag 1500
QY      1570 cagctccgaggcgtcgtgcttaa 1593
        ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||
Db       1501 cagctccgaggcgtcgtgcttaa 1524

RESULT      5
AAV17567
ID          AAV17567 standard; CDNA; 1446 BP.
XX
AC          AAV17567;
XX
DT          10-JUN-1998 (first entry)
XX
DE          Coding sequence for Gy3 subunit of glycinin.
XX
KW          Beta-conglycinin; soybean seed protein; transgenic plant; glycinin;
KM          seed storage protein profile; ss.
XX
OS          Glycine max.
XX
PN          MO97A7731-A2.
XX
PD          18-DEC-1997.
XX
PE          10-JUN-1997; 97MO-U0509743.
PR          14-JUN-1996; 96US-0019940.
PA          (DUPO ) DU PONT DE NEMOURS & CO E. I.
PI          Feder GM, Kinney AJ;
XX
DR          WPI: 1998-052298/05.
XX
PT          Suppression of specific classes of soybean seed protein genes -
PS          useful to change seed storage protein profiles of transgenic plants
XX
PS          Disclosure; Page 36-37; 58pp: English.
CC
CC          This sequence represents the coding sequence for the Gy3 subunit of the
CC          soybean seed protein, glycinin. The method of the invention is for
CC          reducing the quantity of a soybean seed storage protein (A), such as
CC          beta-conglycinin, in soybeans. It comprises: (a) constructing a chimeric
CC          gene comprising: (i) a nucleic acid fragment encoding a promoter that is
CC          functional in the cells of soybean seeds; (ii) a nucleic acid fragment
CC          encoding all or a portion of (A) placed in sense or antisense orientation
CC          relative to the promoter of (i); and (iii) a transcriptional termination
CC          region; (b) creating a transgenic soybean cell by introducing into a
CC          soybean cell the chimeric gene of (a); and (c) growing the transgenic
CC          soybean cells of (b) under conditions that result in expression of the
CC          chimeric gene of (a); where the quantity of one or more members of a
CC          class of (A) subunits is reduced when compared to soybeans not containing
CC          the chimeric gene of (a). The method is used to construct transgenic
CC          soybean lines where the expression of genes encoding (A) are modulated to
CC          effect a change in seed storage protein profile of transgenic plants.
CC          Modification of the seed storage protein profile can result in the
CC          production of novel soy protein products with unique and valuable
CC          functional characteristics.
XX
SQ         Sequence 1446 BP; 426 A; 361 C; 354 G; 305 T; 0 other;.

Query Match      28.5%; Score 527.2; DB 19; Length 1446;
Best Local Similarity 62.2%; Pred.No. 6; 7e-119;
```

Query Match	28.5%;	Score 527.2;	DB 19;	Length 1446;
Best Local Similarity	62.2%;	Pred. No. 6.7e-119;		

Db	1332	tgatcggaactcttgcaggggaacatcatgtttgaacgcattaccagaggaagtgttc	1381
Qy	1487	caaatcatatgtgcctccccaaggagaggaagcaggtctaagaacaacacccctta	1546
Db	1382	agcacacttcaacactaaagaagcagagggcagagcagataaagaacaacacccctta	1441
Qy	1547	agctctcgttcacaccttcacagcagtcctccagagggctgtgactta	1592
Db	1442	agttcctgtgtccactcagcagatcctcagaagaagagagctgtgactta	1487
RESULT 8			
ID	AAQ05359		
XX	AAQ05359 standard; DNA; 1712 BP.		
AC	AAQ05359;		
XX			
DT			
XX	04-DEC-1990 (first entry)		
DE			
XX	Glycinin subunit precursor A2B1a.		
XX			
KW	Glycinin; glycine; pLGVneo1103; ds.		
OS			
XX	Glycine hispida.		
PN	JP02156889-A.		
XX			
PD	15-JUN-1990.		
XX			
PF	08-DEC-1988; 88JP-0310553.		
XX			
PR	08-DEC-1988; 88JP-0310553.		
XX			
PA	(NORQ) NORINSHO.		
XX			
PA	(KIRI) KIRIN BREWERY KK.		
XX			
DR	WPI: 1990-228488/30.		
XX			
PT	Recombinant plasmid - obt'd. by recombining glycinin gene to		
PT	plasmid for plant.		
XX			
PS	Claim 6; Page 561; 20pp; Japanese.		
XX			
CC	Glycinin precursor gene used with exogenous promoter and		
CC	terminator in plasmid pLGVneo1103 to transform plants, improving		
CC	nutritional value.		
XX			
XX			
SO	Sequence 1712 BP; 522 A; 400 C; 414 G; 375 T; 1 other:		
Query Match 27.6%; Score 510.6; DB 11; Length 1712;			
Best Local Similarity 61.7%; Pred. No. 8.1e-115;			
Matches 977; Conservative 0; Mismatches 474; Indels 132; Gaps 5;			
Qy	17	agcttccttttgcctttgcctttcagttctcgtggaagtcagacatctcctcagggcagc	76
Db	52	agcttgccttccttccttgccttcttccttcag--tgcgtgcctgccttgtagagagc	108
Qy	77	agccgagaggaatgcgtgcagttccagccgcctcacaatgcgcagagagacttacacgcga	136
Db	109	agggcagcagaatagagtgccagacaccaaagcctgaaatgcctcacaacaggtaacgcga	168
Qy	137	ttgaatcgaggaagcggtatcatctgagacttgaaccccaacaacagaggtcgaatgcg	196
Db	169	tagagtcggaaggggtcatctgagacttgaaccccaacaacagagcatcagtgctg	228
Qy	197	cggagctgcacctctctcgtctagtcctcgcgcgaacgcctctgtaggcctttact	256
Db	229	ccggtgtgtgcctctctcgtcgtcaccccttaacgcgaatgaccttcgtagacctctaca	288
Qy	257	ccaatgctccccaagagatctcatcagcaggaaggggaactttgggtgatattcc	316

Db	289	ccaaagcgtccccaaggaaactctacatacaacaagaagtatgtatattttgtgcatagtatcc	348
QY	317	ctgtgtgtctctgcacacatactgaagagcctgcgcaacaagaagcgcgatatcagtcacca	376
Db	349	cggtgtgtctctagcattatctcaagagccgcgaagaa	383
QY	377	gaaccacaagaagtttgcagaagaagacccaagccaagccaagaagatagtccaga	436
Db	384	-----tctcagcaacagagacgagacgagagccccaagaccgttcacaa	429
QY	437	aggctgacagcttgcataatgagggttgatctccattgttcagttccaccggtgtgtcttctgc	496
Db	430	aggtaacatcgccttcagagaggggtgattgtgactgcagatgtcctactgtgtctgcattgttga	489
QY	497	tgtaacaagcacaacacacactgtatgtgtgtgtctgttctctactacgtgacaacaacaacg	556
Db	490	tgtaacaacaatgaagaacactcctgtgtgtccgtttctattatgtgacaacaacagcttgg	549
QY	557	acaacagctgtatcagttcccccagggagatccaatttgcgtctggaacacaagacaagagt	616
Db	550	agagacagagctgcacagatgcctagagattctatcttgcctgggaacacaagagcaagagt	609
QY	617	tcttaagtctacaggaacaagaagcagacaagcagagacgaagaagcttacatagcccat	676
Db	610	ttcttaaatatcatcagcaga-----	628
QY	677	acagcccgcatagctgcgccttagacgagagaagcgttgatattgcgccttcgagacagacaa	736
Db	629	-----gcagcaagagagattccc	645
QY	737	gctgcagagagaagcagacagacaagaagaagagacgaagctggaacaatcttcaagcgtc	796
Db	646	aaagccagagaaggaagacaacaagaagaagaacaagagacaacatatgtgaagtgc	705
QY	797	tcaagccgagattctctcgtgacaacagccttccaggttgcagcagacagagattgtgcaaatc	856
Db	706	tgcgccctgtgaattcttgaaagaagacgttctc-----ggtcgtgaacatctgcagatgtgagaaacc	762
QY	857	tgttgaggagagaacagagagtgaagaagaaggagacattgtgaaagtgagagaggagcctcta	916
Db	763	tacaaagtgagaaacgaagagagagatagtgtgagccatctgtgacaagtgaagaagagttcta	822
QY	917	gaattctgagccacgatctggaacgagagagtgctgcgacgaagagagagatacgaattga---ag	973
Db	823	gagtcacaagcctccagccatcagtgagagagccacaagacaagaagaagaatgtatgtgtgagtgatg	882
QY	974	atcaaatatgaatcaccatgaaacagagatgaaagcgtgtgcagagggaaacagagccgggggga	1033
Db	883	agcagcccaagctgtcgtgagacagacaagaagttctgcacaacgcgaagcaaaagagagcagaa	942
QY	1034	atgagattggaagagaagatctgcacccgcagctgtttaaaagaacatttgttgaacaacat	1099
Db	943	atggaattgagaaagacatctgtgcacaatgtagactctgcacaacaatttgttcagaattcat	1002
QY	1094	ccctctacatcatcagacatccctcagccgctgtgttactctcaaacctgcacagcttcaaccttc	1153
Db	1003	cacctgaaatctaaacccccaagcctgtgtagcatctcaacacgcgcacacgccttgccttc	1062
QY	1154	taatccttaagttgctgtgagacttaagtgtctgaatatagtgaaaactctacaggaatgatctg	1213
Db	1063	cagccctctgtcttctcaaaactcagatgtgccagatgtgatactcatctcgcgaagaagtatgt	1122
QY	1214	tgtgcctctacatacaacaacagcagacaagaatcatatagctattgagggagacgggtct	1273
Db	1123	tcgtgcacacatacaacctgtgaacgcgaacacagctaaatatacagctattgaaatgtggcggat	1182
QY	1274	acgtgcagaatgtgtgacagacagcagacaagagtgtgacagcagaggaagcttcaagaggtct	1333
Db	1183	tggtacaagtggtgtgaattgtgaatgtgtgagagaggtgtttgattgtagagagctgcagaaggag	1242
QY	1334	acgttctctgtgtgtgcacagaacttcgcgtgtgtgtgtggaagttccacagacgagaactcgt	1393
Db	1243	gggtgtcgtgatacgttccacaacaaactctgtgcgtgtgtgtgtgcacaaatccacagagcatabattgt	1302

QY	1394	aatcgctggcattcaagaacaggtttcaagggccacgcatgagccaaacttggccggtggaacaact	1453
Db	1303	agatctgtgcatcacaagacacatgatgaaacctcgatctggaacactctgcaggggcaaac	1362
QY	1454	ctctcatagataaccctgcggagaggtgtgtgcaaatctacatgacctccccaaggagagc	1513
Db	1363	cattgttgaacgcatctgcgcagagagtgatctcagacaactttaacttaagaagccagc	1422
QY	1514	aggccaaggcagcgttcaagaacaacacccttcgaagttcttcgtttcaacttttcagcagt	1573
Db	1423	aggccagcgcagtggaagaacaacaacctttagcttccttcttccactcagagagcttc	1482
QY	1574	ctccggaggctgtggtttaaaaa	1596
Db	1483	agaggaagacctgtgcttagaga	1505
RESULT	9		
ID	AAV17566		
XX	AAV17566	standard; cDNA; 1458 BP.	
XX	AAV17566;		
XX	10-JUN-1998	(first entry)	
DE		Coding sequence for Gyz subunit of glycinn.	
XX			
KW	Beta-conglycinin; soybean seed protein; transgenic plant; glycinn;		
KM	seed storage protein profile; ss.		
XX			
OS	Glycine max.		
XX	W09747731-AA2.		
PN	18-DEC-1997.		
XX			
PF	10-JUN-1997;	97MO-US09743.	
XX			
PR	14-JUN-1996;	96US-0019940.	
XX			
PA	(DUPO) DU PONT DE NEMOURS & CO E I.		
XX			
PI	Fader GM, Kinney AJ;		
XX			
PS	WPI: 1998-052298/05.		
PT		Suppression of specific classes of soybean seed protein genes -	
XX		useful to change seed storage protein profiles of transgenic plants	
XX		Disclosure; Page 35-36; 58pp; English.	
XX			
XX		This sequence represents the coding sequence for the Gyz subunit of the	
CC		soybean seed protein, glycinn. The method of the invention is for	
CC		reducing the quantity of a soybean seed storage protein (A), such as	
CC		beta-conglycinin, in soybeans. It comprises: (a) constructing a chimeric	
CC		gene comprising: (i) a nucleic acid fragment encoding a promoter that is	
CC		functional in the cells of soybean seeds; (ii) a nucleic acid fragment	
CC		encoding all or a portion of (A) placed in sense or antisense orientation	
CC		relative to the promoter of (i); and (iii) a transcriptional termination	
CC		region; (b) creating a transgenic soybean cell by introducing into a	
CC		soybean cell the chimeric gene of (a); and (c) growing the transgenic	
CC		soybean cells of (b) under conditions that result in expression of the	
CC		chimeric gene of (a); where the quantity of one or more members of a	
CC		class of (A) subunits is reduced when compared to soybeans not containing	
CC		the chimeric gene of (a). The method is used to construct transgenic	
CC		soybean lines where the expression of genes encoding (A) are modulated to	
CC		effect a change in seed storage protein profile of transgenic plants.	
CC		Modification of the seed storage protein profile can result in the	
CC		production of novel soy protein products with unique and valuable	
CC		functional characteristics.	
XX			
XX	Sequence 1458 BP; 436 A; 348 C; 367 G; 307 T; 0 other;		

[illegible]

[illegible]

RESULT 12

ID	AAQ05357	standard; DNA; 1786 BP.
----	----------	-------------------------

AC AAQ05357;

DT 04-DEC-1990 (first entry)

Glycinin subunit precursor A3B4.

KW Glycine; glycine; PLGvneol103; ds.

05 Glycine hispida.

PN JP02156889-A

~~ad~~ 15-JUN-1990.

XX	08-DEC-1988;	88JP-0310553.
PF		
XX	08-DEC-1988;	88JP-0310553.
PR		
XX		
XX	(NORQ)	NORINSHO.
PA	(KIRI)	KIRIN BREWERY KK.
XX		
DR	WPI, 1990-228488/30.	
XX		
PI	Recombinant plasmid - obid. by recombinating glycinn gene to	
PI	plasmid for plant.	
XX		
PS	Claim 4; Page 561; 20pp; Japanese.	
CC	Glycinn precursor gene may be used with exogenous promoter and	
CC	terminator in plasmid pGlcNeo1103 to transform plants, improving	
CC	nutritional value.	
XX		
SQ	Sequence 1786 BP; 524 A; 452 C; 408 G; 402 T; 0 other;	

Query Match	10.9%;	Score 201.8;	DB 11;	Length 1786;
Best Local Similarity	51.6%;	Pred. No. 1.8e-39;		
Matches 701; Conservative	0;	Mismatches 597;	Indels 61;	Gaps 8;

QY 83 aggagatgcgtccagcttccgcgcgcacaaaggcagagactgataaccgatctgaat 142
|| || | || || || || || || || || || || || || || || || || || || ||
Db 132 agttcaacgagcgccaacttaaaccccaacgcgcttggaaccgcaccacgcgcttgagt 191

143 ccgaagggccggtatcatcttgagacttggaaaccccaacacacagaggttcgaaatcgcgcgcg 202
 192 ccgaaggtggtcttatctgaacatggaactctcaaccccctgagctgcattgcgcgcgtg 251

252 tcaatggttccaacgacccctcaaccgcaagcgctccacatcttaacttaactt 311

Db 312 atcccccaatgatcatctcgttcaagggaaggagcaattggaattgcatlccgggat 371

Db 372 gtcccgagacgttttgagaagccacacacacaaatcaagcaga----- 412

Db 413 -----agaggtcgaaggtctcacagcagcaactcacagacagtcacccgaagattc 461

Dd	462	gtcaactcaatgaagaagacgactagtagtattcttcttgggttcccttaccgagactata	521
Y	503	acgaccacgacactgatttttgcattttcttacttaacacccaacgaacgaacac	562

Db	522	acacibgcgaagaaacaggtttgtgcacatgagtcctcttgaacacctcaactlccaatac	581
Oy	563	agcttgcatacagttccccaagagattcaatttgccttgcctggaaaccacagcagaagttcttaa	622

Db 582 agcttgatcgaacccccagagtatttccctgtcgggaaccagatatgacatccg 641
Oy 623 ggtaccagcaacaagcagacaagcagaagaagcttaccatatagccatagcc 682

642 agaccctgcacaa--cagccgcagcagaagagctcgtgtgcgcgaagcagcggtgcaaca 699
Db
683 cgcatagtcgg-cctcagcaggaagaagcgttgatcttcgcctctgagagacagacagccgc 741
OY

D6 /00 ccgacagcaggaggaagaatgtgcagtctgtccagtgtccttcacaacaacttcttagc 759

QY 742 agagaacgagcagggacagagaaagaacgaagaatgtgaaacattttcag---ggcct 797

700 acaaaLLCCLCaacacacgaggaCaacagctgagaaaattcgtctctcagatgacgaag 819

QY	798	caagccggatcttcctcggaacagcgtccacggttgtagaacagaaagtgtgtaaatct	857
Dd	820	gaagagactcgtcacagtggaggaggagcgttacgcgttaccagcccacaatgtgccagaaca	879
QY	858	gtggggcggaacgcgagatgtgaagaagaggagccattgtgacggttgaggggagcctcag	917
Dd	880	aagaagacgaagaagcgaagacgaagcgaagaataatggaacggagcgccttatctccaag	939
QY	918	aactctgagccccagatggaacgagaggtgcc---gacgaagaagaagaatacatgtgaaga	974
Dd	940	acgacaacagccatgtgaaaagcatgaaagatgacgagacgacgaagaagaagaatcacacc	999
QY	975	tcaatatgtaataccatgnaacagatgagagag-----gggtggcagggg	1016
Dd	1000	ttgtctctgtacctccctccacacggaaccaagcagggcccggaacaacaagaaccaogtggag	1059
QY	1017	aagcacaagcggggggaatgagatgtgaagaagacgaatctgcacgcgatgtgttaaagaagaa	1076
Dd	1060	aggatgtcagactagaagaatgggggttgaggaaaataattgcacccatgnaagcttcaagaa	1119
QY	1077	catttgttggaaaacagatcccccctacaatcagaatccctcagcgctggttactcaaaaatcy	1136
Dd	1120	catgtctgcgccttcaacgtgcgtactcttcaacaaccaaaagctggtgcatalgacacct	1179
QY	1137	ccacgacttcaactctctaactctaagtggctgtggaattgtgctgatatgtgaactc	1196
Dd	1180	caacaactctaccactcccccgcacctccgcgaacttgcgaatcgaatcagttgccataatgtgtcct	1239
QY	1197	ctacaggaatgcatgttctgtccctcaactacacacacgaacgaacagcatcataatatagc	1256
Dd	1240	ctacaggaatggaattacttacttccagatttggaacttgaacgagcagatgtgacg-atga	1298
QY	1257	attgaggggagcgggctcacggttcaagttgtgtagcagcaacggcaacagatgttacgaga	1316
Dd	1299	ctcgaaggaaaggaaga--gttagagttgttgaactgtgcgaagggaatgcagtgttcgacgg	1356
QY	1317	ggagcttcaagagaggtcagcttcttctgtgtgcagaagaacttcgcgcgtgcgtggaggaatc	1376
Dd	1357	tgaactaagagagggaacaattgtctaaagtggcgccagaaaccccgcggtggtctgagcaag	1416
QY	1377	ccagagcgaagacttcgaatacgtgtgcatacagaacaga	1415
Dd	1417	gggagaaacaagatitggaatatgtagtgttcaaagacaca	1455
RESULT 13			
AAAN60939	standard; cDNA; 1786 BP.		
ID	AAAN60939 standard; cDNA; 1786 BP.		
XX	AAAN60939;		
AC	16-OCT-1991 (first entry)		
XX	Sequence encoding soybean glycinin A3B4 subunit.		
DE	Soybean protein; glycinin.		
XX	Glycine max.		
OS	Location/Oualifiers		
XX	Key 47..1597		
FH	CDS /*lag= a		
FT	JP61132189-A.		
XX	19-JUN-1986.		
XX	03-DEC-1984; 84JP-0254217.		
PE	03-DEC-1984; 84JP-0254217.		
XX	(NORO) NORIINSHO KK.		
PA			

XX WPI: 1986-200545/31.
DR P-PSDB; AAP61362.
XX
XX
PT Prepn. of soybean messenger RNA - for insertion into cells or
PT microorganisms to produce soybean protein.
XX
XX
PS Example 1; Fig 1; 7pp; Japanese.
XX
XX Sequence derived from mRNA may be used for the expression of the
CC soybean protein by a foreign host.
XX
XX
SQ Sequence 1786 BP; 524 A; 465 C; 395 G; 402 T; 0 other;

Query Match	10.3%;	Score 191.4;	DB 7;	Length 1786;
Best Local Similarity	50.2%;	Pred. No. 6.2e-37;		
Matches 682;	Conservative	0;	Mismatches 616;	Indels 61;
				Gaps 6

[illegible]

[illegible]

